DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 43, ARTICLE 55, PAGES 1607–1650
PUBLISHED 15 DECEMBER 2020
http://www.demographic-research.org/Volumes/Vol43/55/
DOI: 10.4054/DemRes.2020.43.55

Research Article

Smoothing migration intensities with P-TOPALS

Sigurd Dyrting

Contents

Demographic Research: Volume 43, Article 55
Research Article

Smoothing migration intensities with P-TOPALS

Sigurd Dyrting[1]

Abstract

BACKGROUND
Age-specific migration intensities often display irregularities that need to be removed by graduation, but two current methods for doing so, parametric model migration schedules and non-parametric kernel regression, have their limitations.

OBJECTIVE
This paper introduces P-TOPALS, a relational method for smoothing migration data that combines both parametric and non-parametric approaches.

METHODS
I adapt de Beer's TOPALS framework to migration data and combine it with penalised splines to give a method that frees the user from choosing the optimal number and position of knots and that can be solved using linear techniques. I compare this method to smoothing by model migration schedules and kernel regression using one-year and five-year migration probabilities calculated from Australian census data.

RESULTS
I find that P-TOPALS combines the strengths of both student model migration schedules and kernel regression to allow a good estimation of the high-curvature portion of the curve at young adult ages as well as a sensitive modelling of intensities beyond the labour force peak.

CONCLUSIONS
P-TOPALS is a useful framework for incorporating non-parametric elements to improve a model migration schedule fit. It is flexible enough to capture the variety of profiles seen for both interstate and regional migration flows and is naturally suited to small populations where observed probabilities can be highly irregular from one age to the next.

CONTRIBUTION
I demonstrate a new method for migration graduation that brings together the strengths of

[1] Charles Darwin University, Australia. Email: sigurd.dyrting@cdu.edu.au.

both parametric and non-parametric approaches to give a good general-purpose smoother. An implementation of the method is available as an Excel add-in.

1. Introduction

Researchers working on migration and population projections have long been interested in generating smooth curves of age-specific migration intensities using graduation. The desire of policy makers, health and education administrators, utility providers, and town planners for population projections at ever finer spatial scales has meant that practitioners must estimate migration rates for increasingly small populations, using smoothing to generate stable age-specific probabilities from highly irregular observations (Wilson 2010). Smoothing is also an important tool for extrapolating rates to advanced ages or more generally for generating a complete curve from sparse data (Rogers, Little, and Raymer 2010). For comparative research, demographers require smooth schedules to locate key features, such as the age at which migration intensity reaches its maximum value (Bell et al. 2002; Rees et al. 2002).

Age-specific migration intensity follows a persistent pattern related to major life course events (Bernard, Bell, and Charles-Edwards 2014): intensities are high in the first years of life, decreasing steadily with age in a manner that reflects the mobility of a child's parents. At early adult ages they increase rapidly in response to movements related to work opportunities, reaching a peak in the 20s and thereafter declining as careers and families are established, with an occasional secondary peak at retirement ages. This profile was first given a mathematical form by Rogers, Raquillet, and Castro (1978), who assigned exponential functions to childhood, labour force, and retirement peaks and a constant curve to account for migration independent of age, calling the result a Model Migration schedule (MMS). A fourth exponential component was added by Rogers and Castro (1981) and Rogers and Watkins (1987) to capture the occasional sustained increase in migration intensity with age after retirement associated with movements to access aged care, and a fifth component was added by Wilson (2010) to account for the highly age-concentrated migration of young adults entering tertiary education.

Since their introduction, model migration schedules have been used in a range of contexts, with results confirming both the regular features of the age profile of migration and the usefulness of model migration schedules for fitting them (see Raymer and Rogers (2008) and references therein). Fitting an MMS is not an automatic process. Users need to choose which components to include, set good initial values for each parameter, and devise strategies for ensuring they converge to a sensible solution. Bernard and Bell (2015) compared model schedules with two non-parametric smoothing methods using five-year transition data and found that when correctly specified and fitted, model schedules performed better, but otherwise kernel regression and cubic splines were more reliable. In particular, kernel regression had low variance for small populations and low bias for large populations. Even if an MMS is correctly specified and fitted, it does constrain the shape of the components. This can be a strength when trying to infer schedules from incomplete or noisy observations but becomes a drawback when the objective is to investigate deviations from the paradigm. Congdon (2008) investigated Bayesian approaches to migration graduation and concluded that non-parametric models could detect features in the migration data that MMS could not.

In this article I use interstate migration data at three spatial levels to illustrate the changing requirements of smoothing methods as the population at risk of migrating decreases. Data for all interstate moves displays a low level of noise, and the main requirement of a smoothing method is to accurately fit the observed schedule. State-specific migration profiles can show deviations from the MMS paradigm and an increased level of noise. In this case smoothing methods need to be flexible in the range of profiles they can model and be able to account for the age-dependent size of sample error, particularly for advanced ages. Data for sub-state regions can display very high levels of noise, so that, as the population decreases, it becomes progressively difficult to resolve any feature of the schedule other than its overall level. In this case it is not difficult for a smoothing method to fit observed intensities within the limits set by sample noise. Instead there is more emphasis on the plausibility of the smoothed profile.

MMS and kernel regression methods have their strengths, but we will see that each also has inherent limitations. MMS is good at fitting the highly age-concentrated features seen in one-year curves but lacks a sensitive treatment of probabilities beyond the labour force peak. Kernel regression works well when the distribution is well approximated by a polynomial, as it usually is for five-year probabilities, but not over regions of high curvature, which is often seen in one-year probabilities, or when observed probabilities are unstable across ages, as they often are for small populations or advanced ages. The aim of this paper is to propose a new method that combines the strengths of both parametric and non-parametric approaches and that can serve as an additional tool for migration graduation. My approach is to combine de Beer's (2011) relational Tool for Population Analysis using Linear Splines (TOPALS) with Eilers and Marx's (1996) penalised B-splines (P-splines) to estimate a complete curve of migration probabilities and to show how the resultant nonlinear smoothing equations can be solved using only linear techniques.

In the next section, I summarise the smoothing problem, including a review of transition-style migration data. In Section 3, I introduce the P-TOPALS method and demonstrate its solution by iterated linear regressions. In Section 4, I consider the problem of graduating interstate out-migration as an example of a case where sample noise is small, comparing P-TOPALS with kernel regression and MMS and showing how it can be used to correct a parametric fit and incorporate non-polynomial elements into the age profile. In Section 5, I illustrate how it can be used to smooth state-level in- and out-migration, with emphasis on its flexibility in fitting a range of profile shapes. In Sections 6 and 7, I illustrate how it can be used to smooth interstate and intrastate migration profiles at the sub-state level, with emphasis on its performance under conditions of increasing levels of noise.

2. Smoothing migration probabilities

Migration data of transition type consists of observations

$$ {}_nM = \begin{bmatrix} {}_nM_0 \\ \vdots \\ {}_nM_\omega \end{bmatrix} \quad \text{and} \quad N = \begin{bmatrix} N_0 \\ \vdots \\ N_\omega \end{bmatrix} \tag{1} $$

of ${}_nM_x$ movers of age $x + n$ out of an initial population N_x of age x. Migration probabilities conditional on survival in the country (hereafter just probability) are calculated by taking the ratio

$$ {}_n\tilde{m} = \frac{{}_nM}{N}, \tag{2} $$

where here and in the following all matrix operations and functions act elementwise unless stated otherwise. The problem we consider here is where ${}_n\tilde{m}$ is reported in single-year age groups and our objective is to find a vector

$$ {}_nm = \begin{bmatrix} {}_nm_0 \\ \vdots \\ {}_nm_\omega \end{bmatrix} \tag{3} $$

that in some sense fits ${}_n\tilde{m}$ and is smooth. Conceptually we regard the vector ${}_n\tilde{m}$ as consisting of persistent components ${}_nm$ which we seek to extract and transient features we want to remove. Within a model of migration as a random event occurring to a population exposed to the risk of moving, the transient features have their origin in sampling noise, which becomes relatively less important as the population increases and which shows itself as uncorrelated fitting errors from one age to the next.

3. P-TOPALS

De Beer (2011, 2012) first introduced TOPALS as a tool for fitting and projecting fertility and mortality schedules. The approach is motivated by the observation that by expressing age-specific rates as a product of a standard and a spline, spline weights are stabilised, leading to more realistic projections, and the relationship between target and standard is allowed to be more flexible, leading to better fits.

At the national level, where observed rates show the least amount of noise, de Beer (2011, 2012) showed how the spline weights could be solved in a simple and straightforward way by imposing the condition that the fitted curve equals the observed curve at the spline knots. For graduating death rates at the subnational level, Gonzaga and Schmertmann (2016) showed how to determine the spline weights, taking into account potentially high levels of irregularity, by minimising a Poisson log likelihood function. One common criticism of spline models is that their weights are difficult to interpret because they are not directly related to the value of the fitted curve. Gonzaga and Schmertmann (2016) showed how this could be overcome using linear B-splines (de Boor 2001), which have weights that equal the level at the knot.

Choosing the optimal number and position of knots for a spline fit is not straightforward. Typically, a relatively fine grid is used in regions where the function changes rapidly and a coarse grid is used where it changes slowly (de Beer 2011, 2012; Gonzaga and Schmertmann 2016). An alternative is the P-spline approach (Eilers and Marx 1996), where knots form a fine grid and smoothing is controlled by adding a term to the log likelihood function proportional to a measure of roughness. In this paper I combine TOPALS with P-splines to give a method, P-TOPALS, where smoothing is controlled by a single number, the roughness penalty parameter.

In order to apply P-TOPALS to smooth migration intensities, we need a framework that is independent of the interval n. This is achieved by expressing ${}_nm$ in terms of probabilities at one-year intervals m_k

$$ {}_nm_x = 1 - \prod_{x \le k < x+n} (1 - m_k). \tag{4} $$

For intervals greater than one year ($n > 1$) quantities m_k are to be understood as implied one-year probabilities differing from actual probabilities to the extent that there has been either significant change in migration intensities over the n years preceding the census or significant return/repeat migration and mortality over the same period (Rees 1977). In the TOPALS approach we represent m relative to a standard migration curve $\hat{m}$

$$ \log m = \log \hat{m} + B \cdot \theta, \tag{5} $$

where $\hat{m}$ is an $(\omega + 1) \times 1$ vector, B is an $(\omega + 1) \times l$ matrix of B-spline functions arrayed columnwise, θ is an $l \times 1$ vector of spline weights, and $A \cdot B$ denotes matrix multiplication. The number of B-splines, l is determined by the number of knots (de Boor 2001). Standard TOPALS uses linear splines but higher-order polynomials can also be used.

Following Gonzaga and Schmertmann (2016) I determine spline weights θ by maximising the function

$$\mathcal{L}(\theta) = N' \cdot y - \frac{\lambda}{2}\theta' \cdot D_k' \cdot D_k \cdot \theta, \tag{6}$$

where

$$y = {}_n\tilde{m} \log {}_n m - {}_n m \tag{7}$$

and D_k is the k-order $(l - k) \times l$ difference matrix. Here A' is the transpose of matrix A. The first term in Equation (6) is the log likelihood of observing ${}_nM_x$ movers assuming Poisson counts with mean ${}_nm_xN_x$. The second term introduces a penalty proportional to the square of the k^{th} order finite difference of θ.

Gonzaga and Schmertmann (2016) chose $\lambda = 2$ and $k = 1$ and used the second term to stabilise mortality estimates for very small populations. To find a smooth mortality profile they used a small number of knots (ages 0, 1, 10, 20, 40, 70, and 100), which makes solving Equation (6) feasible using standard nonlinear optimisers. Following Eilers and Marx (1996), I handle the question of the optimal position and number of spline knots by assuming a relatively large number and using the penalty as a means of controlling the smoothness of the fit.

With a large number of knots it is no longer feasible to solve Equation (6) using a multidimensional optimiser. Therefore, an alternative solution method is needed. Maximisation of $\mathcal{L}$ leads to a system of nonlinear equations for θ, which can be solved by iterative linear regressions. Given an approximation $\bar{\theta}$, the updated value θ is calculated by solving the linear equation

$$Q(\bar{\theta}) \cdot \theta = b(\bar{\theta}), \tag{8}$$

where

$$Q(\theta) = G'(\theta) \cdot W(\theta) \cdot G(\theta) + \lambda D_k' \cdot D_k \tag{9}$$

$$b(\theta) = G'(\theta) \cdot V \cdot ({}_n\tilde{m} - {}_n m) + G'(\theta) \cdot W(\theta) \cdot G(\theta) \cdot \theta \tag{10}$$

and

$$W(\theta) = \text{diag}({}_n m \, N) \quad (11)$$
$$V = \text{diag}(N). \quad (12)$$

The derivation of this iteration and the expression for $G(\theta)$ are given in Appendix A. I start the iteration with the constant vector

$$\theta = \log\left(\frac{1}{n}\frac{\sum_x {}_n\tilde{m}_x}{\sum_x \hat{m}_x}\right). \quad (13)$$

3.1 Choosing the penalty

There are a number of criteria for choosing the penalty λ that gives the optimal smoothness (Eilers and Marx 1996). One popular method is Schwarz's (1978) Bayesian information criterion (BIC): λ is found by minimising the function

$$\text{BIC}(\lambda) = -2N' \cdot y + \dim(\theta, \lambda) \times \log(1 + \omega), \quad (14)$$

where the first term is the deviance of the fit (plus a constant) and

$$\dim(\theta, \lambda) = \text{tr}(H) \quad (15)$$

is the effective dimension of θ calculated using the trace of the hat matrix of the linearised problem

$$H = (G' \cdot W \cdot G + \lambda D_k' \cdot D_k)^{-1} \cdot G' \cdot W \cdot G \quad (16)$$

and A^{-1} denotes the matrix inverse of A. Occasionally BIC can give a penalty that is too large, in which case a good alternative is Akaike's (1974) information criterion (AIC): λ is found by minimising the function

$$\text{AIC}(\lambda) = -2N' \cdot y + 2\dim(\theta, \lambda). \quad (17)$$

Criteria such as Equations (14) and (17) seek to find the optimal trade-off between a small deviance and a small dimension. If λ is zero, the deviance will be at its smallest value but the effective dimension will equal its greatest value (the number of knots). As λ

increases, the effective dimension decreases to its minimum value k but the deviance will increase to its maximum value because there are fewer fitting parameters. As a function of the population N, we can say that, all else being equal, the optimal λ will tend to 0 as N increases and will become large as N decreases.

3.2 The role of the P-spline

One of the strengths of model migration schedules is that properly calibrated, they are guaranteed to give sensible age profiles. One of their weaknesses is their parametric nature, which imposes limits on their fidelity. The P-TOPALS framework can be used as a means of improving a fit by including non-parametric elements. To illustrate this, consider the special case of one-year probabilities ($n = 1$). Let $\hat{m}$ be an MMS fit to an observed migration profile. The quality of the fit can be judged by examining the residuals

$$r = \log \tilde{m} - \log \hat{m} \tag{18}$$

for structure. For example, when they used standard MMS to fit Chilean inter-provincial and inter-municipal migration probabilities, Bernard and Bell (2015) found that residuals had a persistent and strong age profile and positive auto-correlation. Our objective is then to find an improved fit m such that the new errors

$$\epsilon = \log \tilde{m} - \log m \tag{19}$$

are uncorrelated. Substituting Equation (18) and Equation (19) into Equation (5) and rearranging gives the relation

$$r = B \cdot \theta + \epsilon, \tag{20}$$

which shows that the role of the P-spline is to fit the residuals. Equation (6) tells us that using P-TOPALS will never lead to a worse fit in the sense that the weights θ will only be non-zero if m gives a greater log likelihood than $\hat{m}$.

3.3 The role of the standard

Expressing m in the form Equation (5) is convenient for projecting rates because convergence to a standard can be modelled by letting $\theta \to 0$ over time (de Beer 2011, 2012), but it is also useful for reducing the number of knots necessary to fit a schedule. The reason is that polynomial approximations struggle in regions close to either a vertical asymptote (the first year of life for mortality) or a horizontal asymptote (near age 15 or 50 for fertility). Expressing m in the form Equation (5) allows us to effectively remove these elements from the problem by packing them into the standard $\hat{m}$. This is also the reason why for smoothing purposes the choice of the standard is not that important when the population is reasonably large (provided it includes the non-polynomial parts of the schedule), as has been observed by both de Beer (2011) and Gonzaga and Schmertmann (2016).

The role of the standard can be made more precise by considering the two limits of a small and large penalty. When the optimal penalty is chosen using one of the information-based criteria of Section 3.1, these two cases correspond to the large N and small N limits. When λ is small, the case used by both de Beer (2011, 2012) ($\lambda = 0$) and Gonzaga and Schmertmann (2016) ($\lambda = 2$), the first term on the right-hand side of Equation (6) dominates. In this case two standards $\hat{m}_1$ and $\hat{m}_2$ that differ precisely by a B-spline

$$\log \hat{m}_1 = \log \hat{m}_2 + B \cdot \phi_{12} \tag{21}$$

will give identical fitted curves m with spline weights related by

$$\theta_1 = \theta_2 - \phi_{12}. \tag{22}$$

In other words, B-spline deformations applied to the standard will have no effect on the fitted curve. A corollary to this result is that for smoothing in the presence of a small penalty, the role of the standard is to model those portions of the age distribution that are not well represented by a B-spline – that is, those parts that are not locally polynomial. We will see that for one-year migration probabilities, these are ages where the change in level is effectively discontinuous, and for multi-year probabilities, these are ages where the change in slope is discontinuous.

When λ is large, the second term on the right-hand side of Equation (6) will dominate. For the case $k = 1$, this leads to the solution

$$\theta = \iota\theta_0, \tag{23}$$

where θ_0 is a free parameter and ι is a vector of ones. Since B-splines form a partition of unity, that is $B \cdot \iota = 1$, it follows that

$$m = \hat{m}e^{\theta_0}, \tag{24}$$

which shows that in this case, the role of the standard is to determine the entire profile of migration probabilities up to a multiplicative constant. We will see that this property works to stabilise fits to data from small populations.

4. Application to interstate migration

Interstate migration probability is a measure of national internal mobility obtained by dividing the number of people who have moved interstate over a specified period by the total population, movers and non-movers. It is an ideal test case for graduation methods because it samples the entire population and is therefore most free of the confounding effect of noisy data.

Migration data by state and single year of age over one- and five-year intervals was obtained from the Australian Bureau of Statistics (ABS) 2016 Census of Population and Housing and used to calculate raw interstate migration probabilities out to age $\omega = 90$. The results are shown in Figure 1, together with curves obtained using kernel regression and MMS. For the kernel regression fit I chose local linear polynomials and a Gaussian kernel (Fan and Gijbels 1996). The kernel bandwidth was calculated using Ruppert, Sheather, and Wand's (1995) rule-of-thumb plug-in bandwidth selector. For the MMS fit I chose Wilson's (2010) 16-parameter student model

$$\begin{aligned} m_x \quad = \quad & a_1 \exp(-\alpha_1 x) && \textit{(childhood)} \\ & +a_2 \exp\left(-\alpha_2(x-\mu_2) - e^{-\lambda_2(x-\mu_2)}\right) && \textit{(labour force)} \\ & +a_3 \exp\left(-\left(\tfrac{(x-\mu_3)}{\sigma_3}\right)^2\right) && \textit{(retirement)} \\ & +a_4 \exp\left(\alpha_4 x\right) && \textit{(elderly)} \\ & +a_5 \exp\left(-\alpha_5(x-\mu_5) - e^{-\lambda_5(x-\mu_5)}\right) && \textit{(student)} \\ & +c && \textit{(constant)} \end{aligned} \qquad (25)$$

because Australian interstate migration over a one-year interval has a well-defined student migration peak. I set the elderly component to zero because the data does not exhibit a post-retirement increase in migration intensity. An initial guess for the remaining 14 parameters was refined using the sequential method described in Wilson (2010) and then a final polishing of the values was done by minimising the sum of squared errors using a nonlinear optimiser.

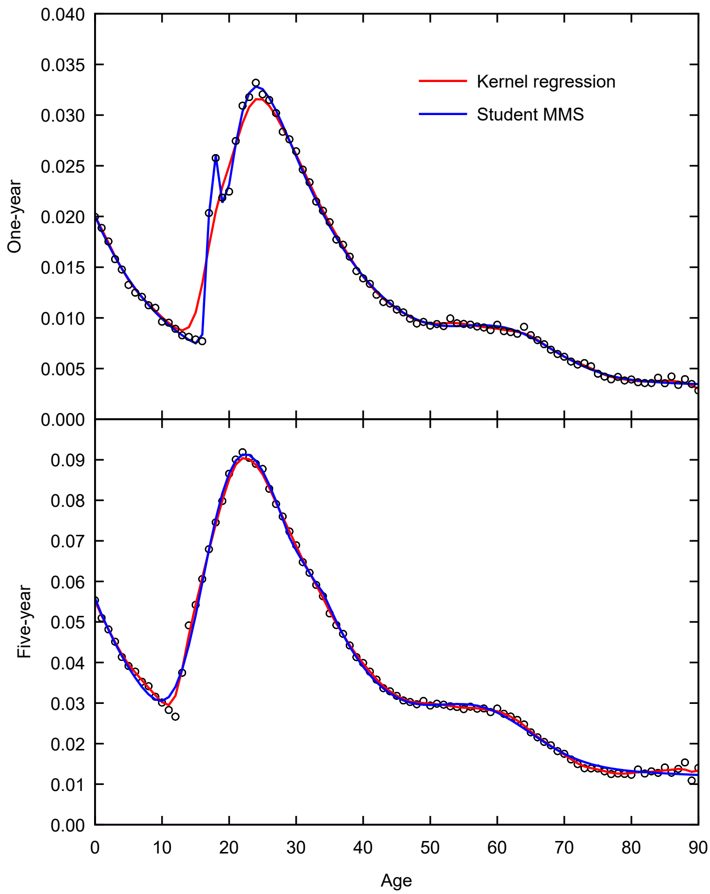
The top panel of Figure 1 shows fits to one-year data. We see that MMS performs better than kernel regression over the 15–25 age range. It captures the sudden jump from age 16 to 17 and the minor peak at age 18, whereas kernel regression gives a more gradual increase from age 14 and a monotonic increase in level that peaks at age 24. Kernel regression is a bad fit for these ages mainly because it assumes migration is well approximated locally by a polynomial function of age when in fact the change in probability from 16 to 17 is effectively discontinuous. A second reason is that the rule-of-thumb bandwidth selector calculates a constant bandwidth that is applied to all ages. Automatic variable bandwidth selectors have been proposed (Fan and Gijbels 1996), but implementations are not widely available at this time.

What is perhaps not so clear is that after age 30, kernel regression provides a better fit than MMS. This can be seen in Figure 3, which plots the cumulative sum of squared errors from age 30 in the top panel and age-specific standardised residuals in the bottom panel.[1] Comparing the slopes of the cumulative sum of squared error curves and the amplitudes of the standardised residuals we see that for MMS errors are accumulated at a rate greater than for kernel regression. The relatively poor performance of MMS, over this age range is probably due to limitations imposed by a parametric profile. Thus we see that for one-year interstate migration, neither method can be preferred over the entire age range. As discussed in Section 3.2, P-TOPALS can be used to improve an MMS fit. In this case I take as the standard $\hat{m}$ the student MMS fit, given in the top panel of Figure 1. For the P-TOPALS fit I used linear basis splines with knots spaced three years apart from ages 0 to 90 and a linear penalty ($k = 1$) with the penalty determined by the BIC condition. The top panel in Figure 2 shows the P-TOPALS fit. The fit to student peak has been preserved and as Figure 3 shows, after age 30 the amplitude of standardised errors has been reduced and the cumulative sum of squared errors grows at the same rate as for kernel regression.

The bottom panel of Figure 1 shows fits to five-year data. Comparing one- and five-year data we see that the five-year age profile is smoother. In particular the increase in migration intensity leading to the labour force peak is more gradual than for one-year migration intensities. As a result, the kernel regression fit has improved over these ages. Figure 4 shows the cumulative sum of squared errors, in the top panel and standardised residuals in the bottom panel. We see that near age 12, both kernel regression and MMS have a large spike in standardised residual and a consequent jump in cumulative sum of squared errors but the size is smaller for kernel regression. Beyond age 30, comparing the slopes of the cumulative sum of squared error curves and the amplitudes of the standardised residuals we see that for MMS, errors are accumulated at a rate greater than for kernel regression.

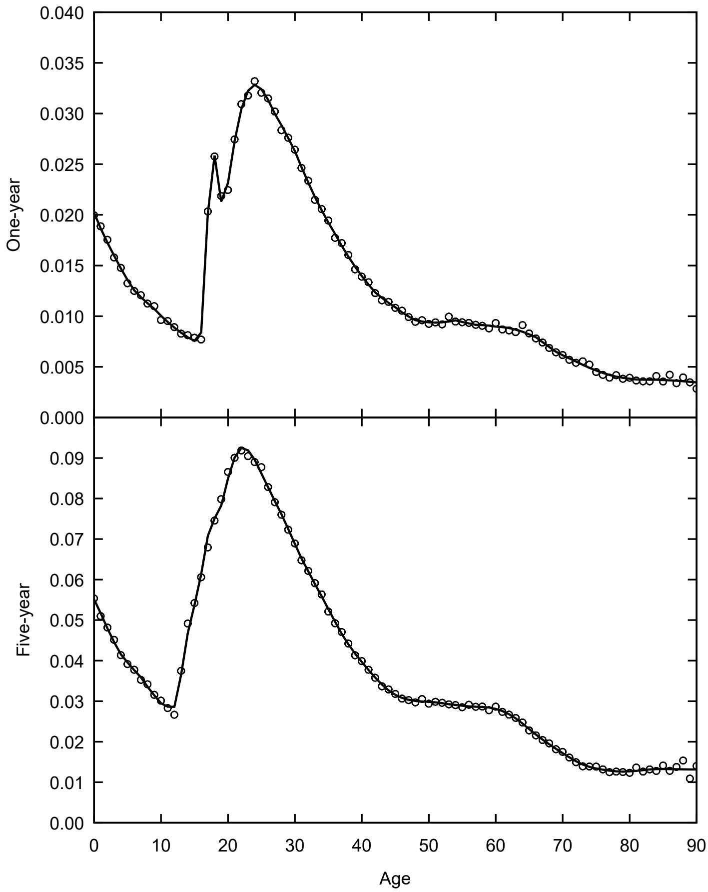
[1] Residuals are standardised by dividing by the Poisson standard deviation $\sqrt{{}_n m/N}$.

Figure 1: Australian interstate migration probabilities 2016, two smoothing methods



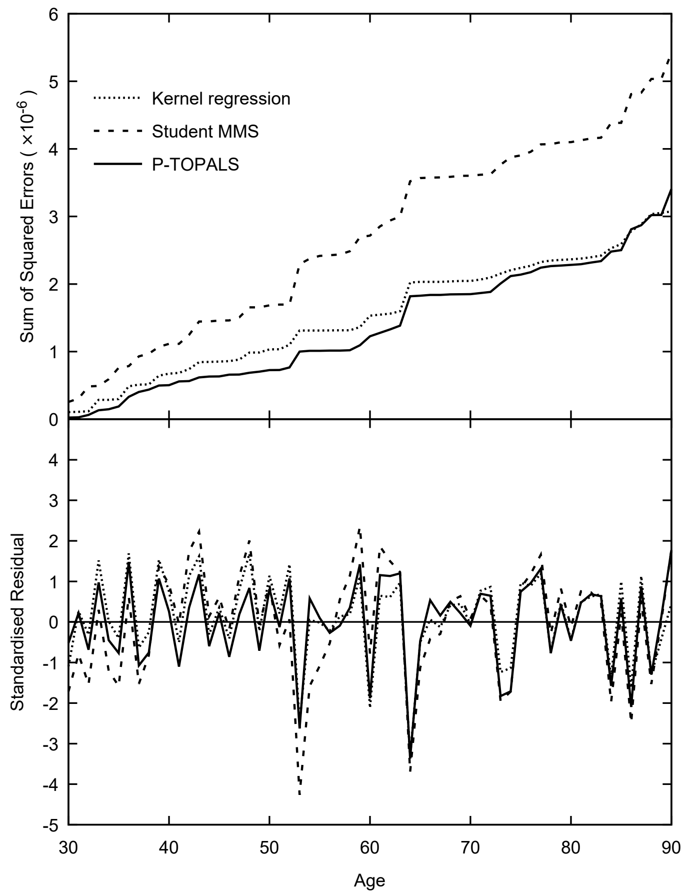
Note: Age is in completed years at the beginning of the migration interval.
Source: Based on ABS data.

Figure 2: Australian interstate migration probabilities 2016, smoothing with P-TOPALS



Note: Age is in completed years at the beginning of the migration interval.
Source: Based on ABS data.

Figure 3: Australian interstate migration 2015–2016, smoothing errors from age 30



Note: Age is in completed years at the beginning of the migration interval.
Source: Based on ABS data.

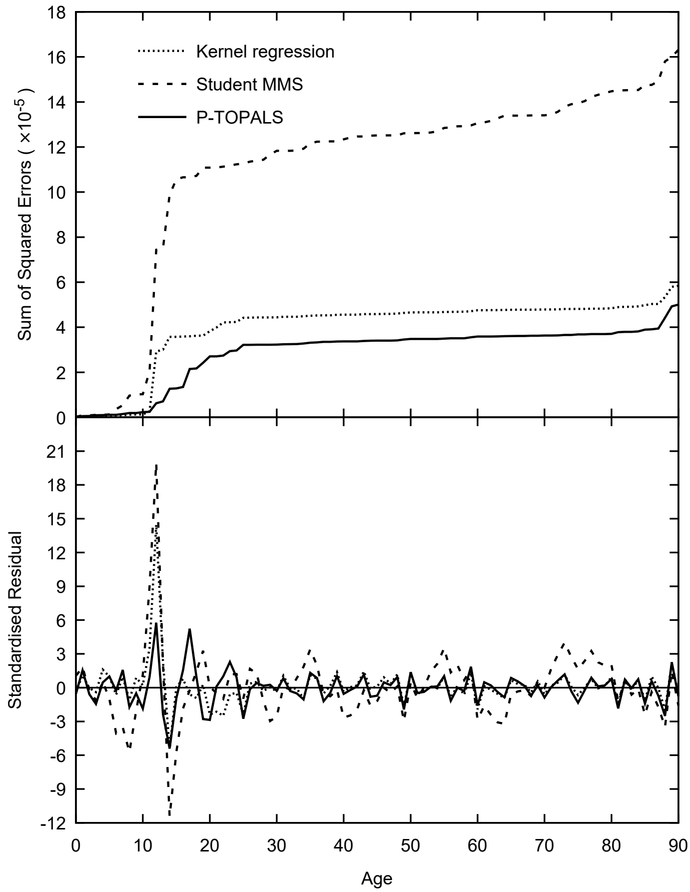
For multi-year probabilities, life course events are imprinted not only on the level of the age profile but also on its slope. For example, raw five-year probabilities in Figure 1 display a sudden change in slope at age 12 which suggests the existence of a student peak in the implied one-year probabilities. This can be seen by expanding Equation (4) to terms first order in m_k and taking the difference to get

$$\Delta_n m_x = {}_n m_{x+1} - {}_n m_x \approx m_{x+n} - m_x, \tag{26}$$

which shows that a sudden increase in the slope of the five-year probability at age 12, indicates a jump in the implied one-year probability at age 17. Both kernel regression and MMS are over-smoothing near age 12, which is clearly demonstrated in Figure 4 by the sudden increase in the cumulative sum of squared errors and the large spike in standardised residual that both methods display at this age. This is to be expected for kernel regression because of its local polynomial assumption. In the case of MMS, this is occurring because the choice of the square of absolute errors as a fitting metric tends to favour fitting for ages where migration probability is highest, whereas the feature we are trying to fit occurs over a small number of points at a low level. There are options for improving the MMS fit over the student years. Changing the error metric from absolute to relative errors worked for the 2006 census data but not for 2011. Increasing the weighting of this part of the objective function relative to other ages gave good results, although the fit after the labour peak became worse.

Section 3.3 showed how P-TOPALS can be used to add non-polynomial elements to a fit. In this case I take as the standard $\hat{m}$ the one-year P-TOPALS fit in the upper panel of Figure 2, which has a jump in migration intensity at age 17. The bottom panel shows the P-TOPALS fit to five-year data. We see that P-TOPALS is able to capture the sudden change in slope, and as a result, Figure 4 shows a reduction in the spike in standardised residuals at age 12 and a more gradual increase in the sum of squared errors over, ages 10 to 20. For ages 30 and over, the size of the standardised residuals and the slope of the cumulative sum squared errors are very close to kernel regression.

Figure 4: Australian interstate migration 2011–2016, smoothing errors



Note: Age is in completed years at the beginning of the migration interval.
Source: Based on ABS data.

5. Application to individual states

State in- and out-migration probabilities are measures of sub-national mobility obtained by dividing the number of people who have moved to and from a state by the population outside and within the state, respectively.

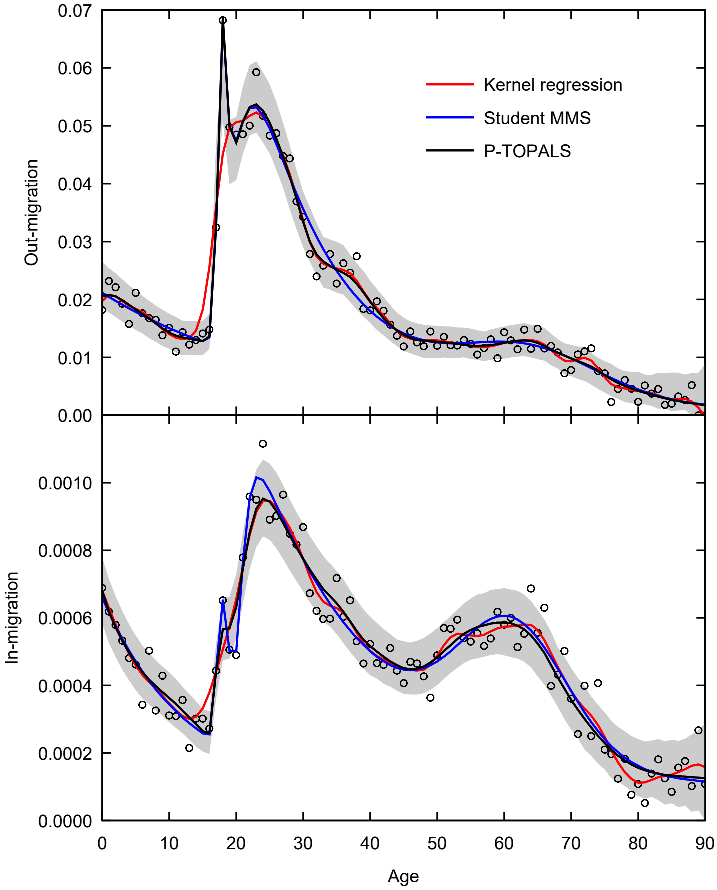
Data from the 2016 Australian Census of Population and Housing was used to calculate raw age-specific in-migration and out-migration schedules for each of Australia's six states and two mainland territories over a one-year interval. The observed profiles for Tasmania are shown in Figure 5, together with smoothed curves obtained using kernel regression, student MMS, and P-TOPALS. Also shown is the 95% confidence interval for observed intensities based on the P-TOPALS fit. For reasons of space, figures for the other seven states and territories are given in Appendix B. Since one state's departure is another's arrival, it follows that interstate migration is a weighted average of in-migration or out-migration, where the weights are the population outside (in-migration) or within (out-migration) a state as a fraction of the total. However, comparing the top panel of Figure 1 with Figures 5, and B-1 to B-7, we see that there can be considerable deviation from this average.

Table 1: Summary statistics for three smoothing methods applied to Australian state and territory out- and in-migration probabilities, 2015–2016

State	Out/In	dev_0			dev_{30}			Notes			Fig.
		K	M	P	K	M	P	K	M	P	
NSW	Out	837	219	91	31	105	36	S	L		B-1
	In	237	116	92	43	84	65	S			
VIC	Out	237	108	108	46	86	64	S			B-2
	In	529	111	104	61	76	73	S			
QLD	Out	299	125	97	55	92	72	S			B-3
	In	425	159	116	66	106	72	S	L		
WA	Out	265	149	141	66	97	87	S			B-4
	In	129	207	154	71	125	88		L		
SA	Out	221	191	112	67	153	78	S	L		B-5
	In	174	139	129	74	104	92	S			
TAS	Out	174	126	95	55	99	73	S	L		5
	In	150	145	163	75	108	105	S,E		S	
ACT	Out	263	152	159	99	118	123	S			B-6
	In	926	212	181	105	149	120	S	L		
NT	Out	188	146	139	78	97	98	S,E			B-7
	In	97	182	144	71	139	109		L		

Note: Goodness-of-fit measures dev_0 and dev_{30} are Poisson deviances for ages greater than or equal to $a = 0$ and $a = 30$, respectively, given by Equation (27). K, kernel regression; M, student model migration schedule for all flows except WA and NT in-migration, which use the standard model migration schedule, Equation (B-1); P, P-TOPALS. The "Notes" column gives the author's assessment of a fit's deficiencies, if any: S, over-smoothing the student peak; L, over-smoothing the profile after the labour peak; E, under-smoothing advanced ages.

Figure 5: Tasmania one-year migration probabilities 2015–2016 by age, three smoothing methods



Note: Age is in completed years at the beginning of the migration interval. Grey area, 95% confidence interval for observed intensities based on P-TOPALS fit.
Source: Based on ABS data.

Appendix B summarises the strategies used to fit the observed probabilities with our three methods. Table 1 gives two summary statistics of fidelity of each fit: a measure of the global fit, dev_0, and a measure of the fit after the labour force peak, dev_{30}, where the variable dev_a, given by the expression

$$\mathrm{dev}_a = 2 \sum_{x \geq a} N_x \left[{}_n\tilde{m}_x \log \left({}_n\tilde{m}_x / {}_n m_x \right) - \left({}_n\tilde{m}_x - {}_n m_x \right) \right], \tag{27}$$

is the Poisson deviance of the fit for ages a and over. Of the 16 schedules fitted, kernel regression had the lowest value for dev_0 for two cases (in-migration for Western Australian and the Northern Territory), MMS had the lowest value for three cases (out-migration for Victoria and the ACT, and in-migration for Tasmania), and P-TOPALS had the lowest value for 12 cases. For each of the 16 schedules, kernel regression had the lowest dev_{30}, and except for Northern Territory out-migration, MMS had the highest dev_{30}.

In Table 1 I also give my assessment of each fit's deficiencies, if any, focussing on three types: over-smoothing of the student peak, over-smoothing of the profile after the labour force peak, and, for kernel regression, under-smoothing of the profile at ages 80 and over. Assessment was first made graphically and then checked against the measures of global fit dev_0 and post-labour force fit dev_{30}.

Over-smoothing of the student peak or post-labour peak corresponded to large values of dev_0 and dev_{30}, respectively. In general we see that student MMS provides a good fit before the labour force peak but not always after it (see out-migration in Figures 5, B-1 and B-5, and in-migration in Figures B-3, B-4, B-6, and B-7). Kernel regression over-smooths the student peak whenever it is present (see out-migration in Figures 5 and B-1 to B-7 and in-migration in Figures 5, B-1 to B-3, B-5, and B-6) but in general provides good fits after it, with the smallest values of dev_{30}. P-TOPALS gives good fits after the labour force peak, showing remarkably similar age profiles to kernel regression despite the two methods being based on different algorithms (see out-migration in Figures 5, B-1, and B-5 and in-migration in Figures B-3, B-4, B-6, and B-7). The quality of the P-TOPALS fit before the labour force peak depends on whether a student peak is present and if so the choice of the standard. In its fit to Tasmania in-migration (Figure 5) we see that the method is over-smoothing the student peak. In this case the standard was the student MMS fit to in-migration aggregated over 2006, 2011, and 2016 censuses, which had a less pronounced peak at age 18 than the 2016 curve.

There is an increase in the level of irregularity in observed migration rates as population N decreases from the larger states (see Figures B-1 to B-3) to the midsize (see Figures B-4 and B-5) and the smaller ones (see Figures 5, B-6, and B-7). Student MMS and P-TOPALS are robust but kernel regression can have problems smoothing for advanced ages where the population at risk of migrating is small and observed intensities are highly irregular. A symptom of this is the appearance of oscillations in the fitted age profile for ages over 60 (see out-migration in Figure B-7 and in-migration in Figure 5), with both cases corresponding to low values of dev_{30} when student MMS and P-TOPALS values are close.

As discussed in Section 3.3, when N is sufficiently large, the P-TOPALS standard will not be required for smoothing if the age-specific migration intensities do not exhibit non-polynomial features. As an illustration of this I have used a flat standard $\hat{m} = 1$ in the P-TOPALS fit to Western Australian and Northern Territory in-migration, which neither have a strong student peak nor show excessive levels of sample noise (Figures B-4 and B-7). In these two cases we see that the P-TOPALS fit is similar to the kernel regression profile.

6. Application at the sub-state level

Interstate out-migration at the sub-state level is obtained by dividing the number of people from a sub-state area who have moved interstate by the population in the area. Data from the 2016 Australian Census of Population and Housing was used to calculate raw age-specific interstate out-migration schedules for four areas within the Northern Territory spanning three sub-state geographic divisions (Australian Bureau of Statistics 2016): Darwin (SA4), Darwin Suburbs (SA3), Nightcliff (SA2), and Moil (SA2). From the observed profiles for one-year and five-year intervals shown in Figures 6 to 9 we see that methods for smoothing migration probabilities for populations at the sub-state level must adapt to an increasing noise amplitude as the population size decreases and a changing requirement from smoothing irregularities to imposing regularities (Rogers, Little, and Raymer 2010).

The ABS, like other national statistical offices, provides a platform for the customised tabulation of census data by remote server as part of a program to increase the access and use of official statistics (Australian Bureau of Statistics 2019). The requirement that publicly available data be non-disclosive means that irregularities in sub-state migration probabilities derived from such tables are due to two sources of noise: the random nature of migration events and perturbations added to table cells as part of a confidentialisation procedure (Thompson, Broadfoot, and Elazar 2013). Of the two numbers used to calculate migration probability in Equation (2), perturbation has the greatest relative impact on the migration flow ${}_nM$ because perturbation variance is a bounded function of cell magnitude (Fraser and Wooton 2005; Wooton 2006; Australian Bureau of Statistics 2019). Neglecting the effect of perturbation on N and assuming that actual migration is a Poisson process, it follows that the variance in estimates of migration probability derived from perturbed tables is given by

$$\mathrm{Var}\left({}_n\tilde{m}\right) \approx \frac{{}_nm}{N} + \frac{V_p}{N^2}, \tag{28}$$

where V_p is the perturbation variance. Equation (28) shows that perturbation noise increases more rapidly than sample noise as N decreases and becomes the same order of size as sample noise when the population per age group approaches the value

$$N_p = 100\frac{V_p}{\mathrm{CMP}}, \tag{29}$$

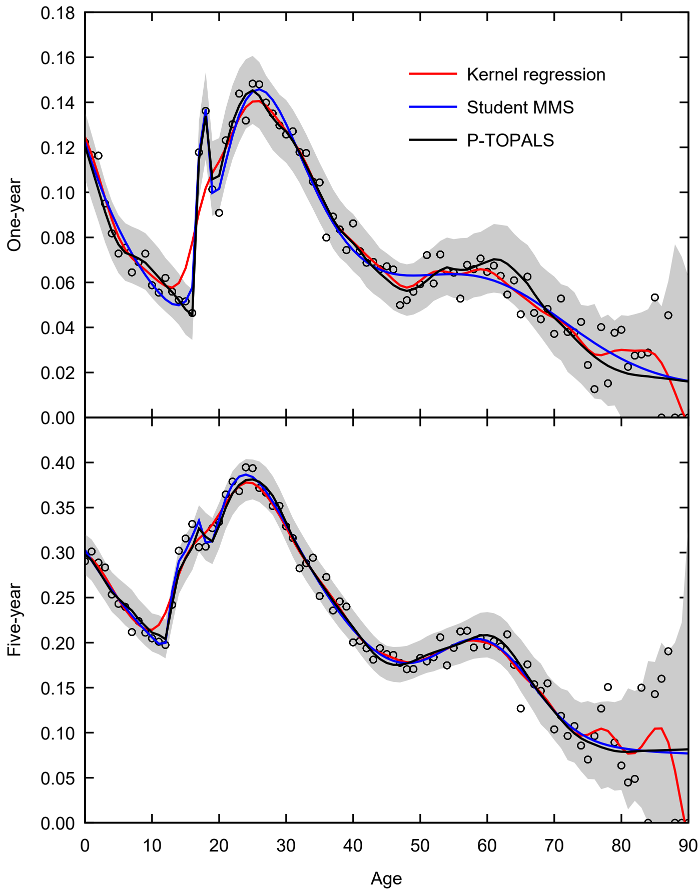
where CMP is the crude migration probability (Bell et al. 2002). The value of V_p can be estimated by comparing migration flows aggregated by five-year age groups with the sum of migration flows by single year, from which I found that for one-year migration intervals, $V_p \approx 2.3$. The one-year crude interstate out-migration probability for the Northern Territory is approximately $\mathrm{CMP} \approx 7.8$, which gives $N_p \approx 30$. For Nightcliff, N is below this value for ages above 60. For Moil, N is below or close to this value for all ages. For five-year migration intervals $V_p \approx 3.2$. The five-year crude interstate out-migration probability for the Northern Territory is approximately $\mathrm{CMP} \approx 23.0$, which gives $N_p \approx 14$. For Nightcliff, N is below this value for ages above 70. For Moil, N is below this value for ages above 60.

The four sub-state areas, Darwin, Darwin Suburbs, Nightcliff, and Moil, form a hierarchy, each enclosed by an area of higher order in the Australian Statistical Geography Standard (NT, Darwin, and Darwin Suburbs, respectively). In the kernel regression fits I used the same configuration as for Section 5: linear polynomials and a Gaussian kernel with a global bandwidth calculated using the rule-of-thumb method. In the student MMS fits I used the fitted parameter values of the enclosing area as starting values. Similarly, for the P-TOPALS fits I used the fitted schedule of the enclosing area as the standard and quadratic basis splines with penalty chosen using BIC. To account for perturbation noise in the data for Nightcliff and Moil I used the adjusted exposure

$$N^* = \frac{N}{1 + N_p/N}, \tag{30}$$

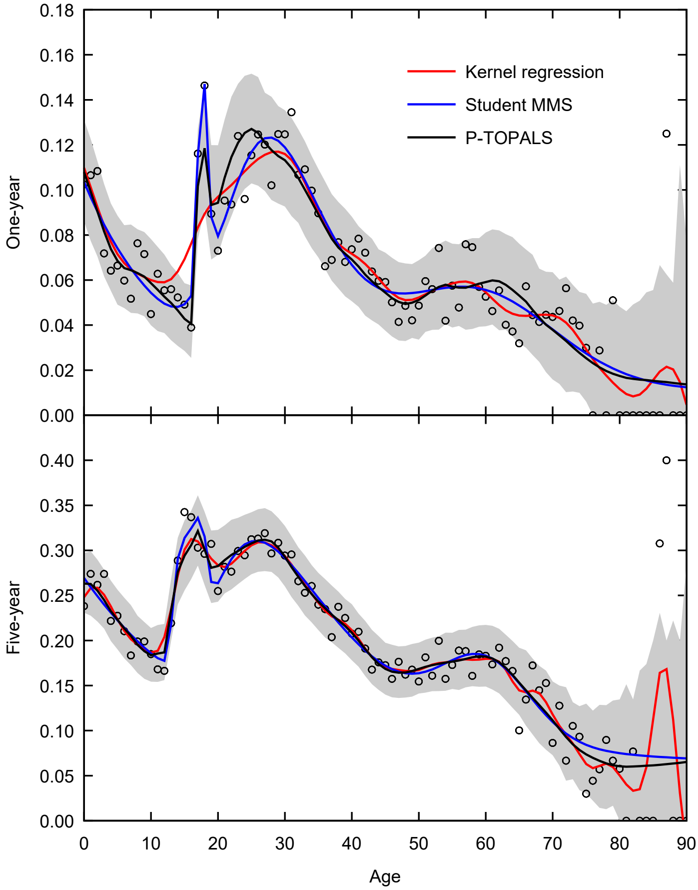
which follows from the condition that the adjusted variance ${}_nm/N^*$ should approximate the expression Equation (28).

Figure 6: Darwin interstate out-migration probabilities 2016 by age, three smoothing methods



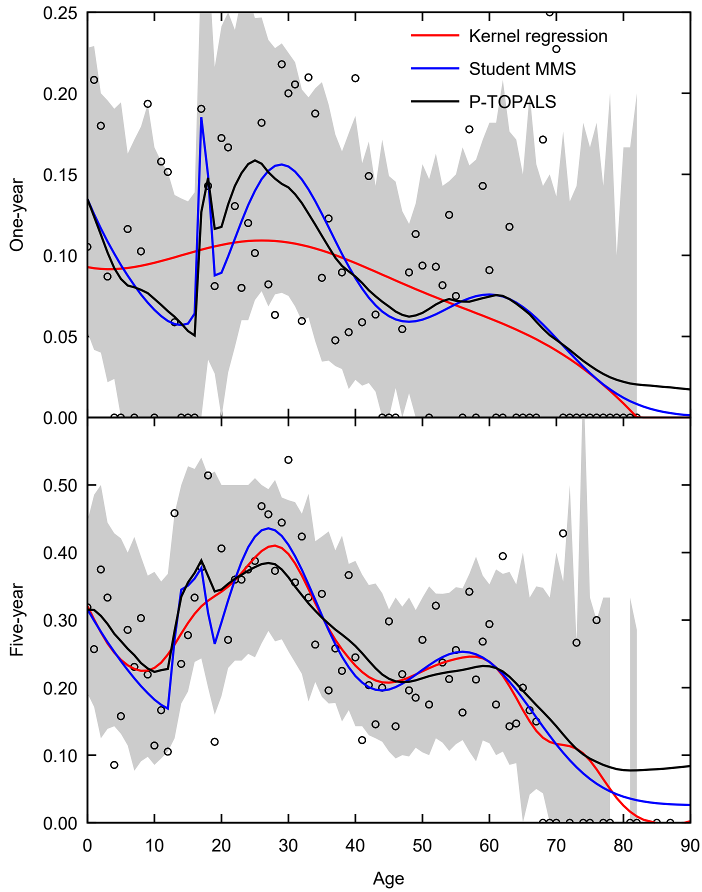
Note: Age is in completed years at the beginning of the migration interval. Grey area, 95% confidence interval for observed intensities based on P-TOPALS fit.
Source: Based on ABS data.

Figure 7: Darwin Suburbs interstate out-migration probabilities 2016 by age, three smoothing methods



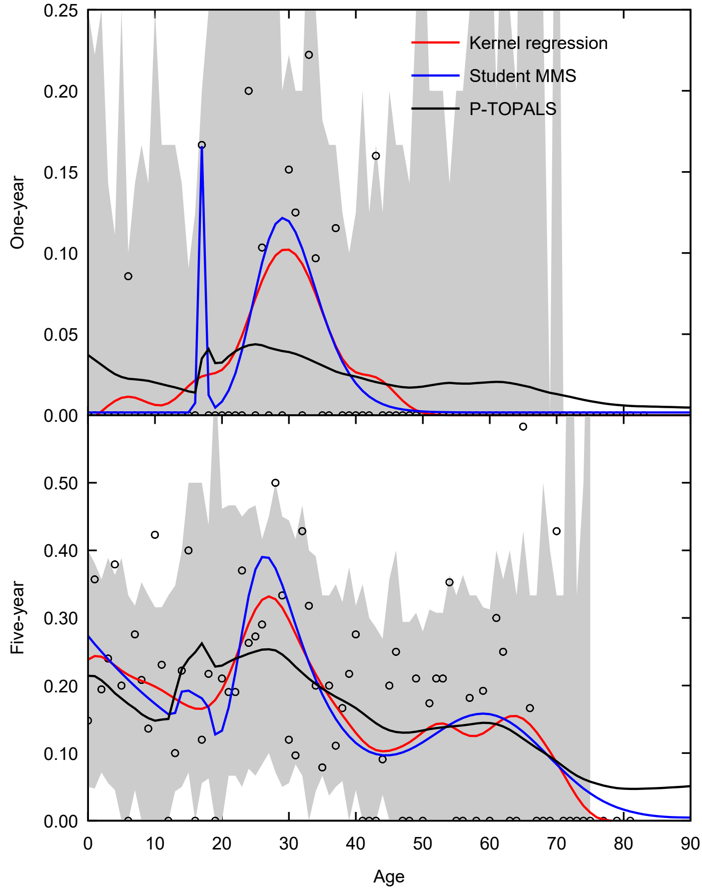
Note: Age is in completed years at the beginning of the migration interval. Grey area, 95% confidence interval for observed intensities based on P-TOPALS fit.
Source: Based on ABS data.

Figure 8: Nightcliff interstate out-migration probabilities 2016 by age, three smoothing methods



Note: Age is in completed years at the beginning of the migration interval. Grey area, 95% confidence interval for observed intensities based on P-TOPALS fit.
Source: Based on ABS data.

Figure 9: Moil interstate out-migration probabilities 2016 by age, three smoothing methods



Note: Age is in completed years at the beginning of the migration interval. Grey area, 95% confidence interval for observed intensities based on P-TOPALS fit.
Source: Based on ABS data.

Table 2 shows the average population per single year of age $\bar{N}$ as well as two summary statistics for each smoothing method: a measure of global goodness-of-fit dev_0 and a measure of plausibility-of-shape $\bar{P}$. As $\bar{N}$ decreases it becomes progressively difficult to resolve in the data any feature beyond the overall intensity of migration. Smoothing methods with a fixed or large number of parameters can begin to give unrealistic schedules as these parameters are increasingly used to fit noise. As a measure of shape plausibility I use the percentage of a schedule's profile that differs from a reference profile

$$\bar{P} = 100\left(1 - \frac{m'_{\mathrm{ref}} \cdot m}{|m_{\mathrm{ref}}||m|}\right), \tag{31}$$

where m_{ref} is a reference schedule and $|v|$ is the absolute value of vector v. For reference schedules I used the P-TOPALS smoothed interstate probabilities shown in Figure 2. Also included in Table 2 is my assessment of a fit's deficiencies, if any.

Table 2: Summary statistics for three smoothing methods applied to interstate migration for four sub-state areas, 2016

Area	n	$\bar{N}$	dev_0			$\bar{P}$			**Notes**			**Fig.**
			K	**M**	**P**	**K**	**M**	**P**	**K**	**M**	**P**	
Darwin	1	1,602	107	80	73	2	1	1	S,E			6
Darwin Suburbs	1	627	128	110	116	2	2	1	S,E			7
Nightcliff	1	51	219	195	196	6	3	1	X			8
Moil	1	25	93	86	132	17	26	1	X	X		9
Darwin	5	1,425	84	86	95	2	2	2	S,E			6
Darwin Suburbs	5	585	66	88	85	3	3	2	E			7
Nightcliff	5	48	89	89	99	3	3	3				8
Moil	5	25	183	185	205	6	5	2				9

Note: Quantities n and $\bar{N}$ are the migration interval in years and the average population per single year of age, respectively. Goodness-of-fit measure dev_0 is the Poisson deviance for all ages given by Equation (27). Quantity $\bar{P}$ measures the percentage of a schedule's profile that differs from the interstate profile given in the top panel ($n = 1$) and bottom panel ($n = 5$) of Figure 2. K, kernel regression; M, student model migration schedule; P, P-TOPALS. The "Notes" column gives the author's assessment of a fit's deficiencies, if any: S, over-smoothing the student peak; E, under-smoothing advanced ages; X, implausible shape.

For Darwin and Darwin Suburbs, MMS and P-TOPALS appear to perform equally well, with similar values for dev_0 and $\bar{P}$ for both one- and five-year probabilities. For one-year probabilities kernel regression is over-smoothing the student peak and as a result its dev_0 is somewhat larger. Note that for five-year probabilities it has the smallest value for dev_0 but is clearly under-smoothing the profile for advanced ages. For Nightcliff and Moil, one-year probabilities, kernel regression and MMS fits have increased values for $\bar{P}$, with both having implausible profiles for Moil despite low values for dev_0. P-TOPALS continues to give plausible shapes and low values for $\bar{P}$, which illustrates the role of the standard in stabilising fits for small N discussed in Section 3.3. For Nightcliff and Moil, five-year probabilities kernel regression and MMS have similar values for dev_0 and $\bar{P}$. P-TOPALS has a higher dev_0 but a low $\bar{P}$ for Moil, where $\bar{P}$ for kernel regression and MMS has increased.

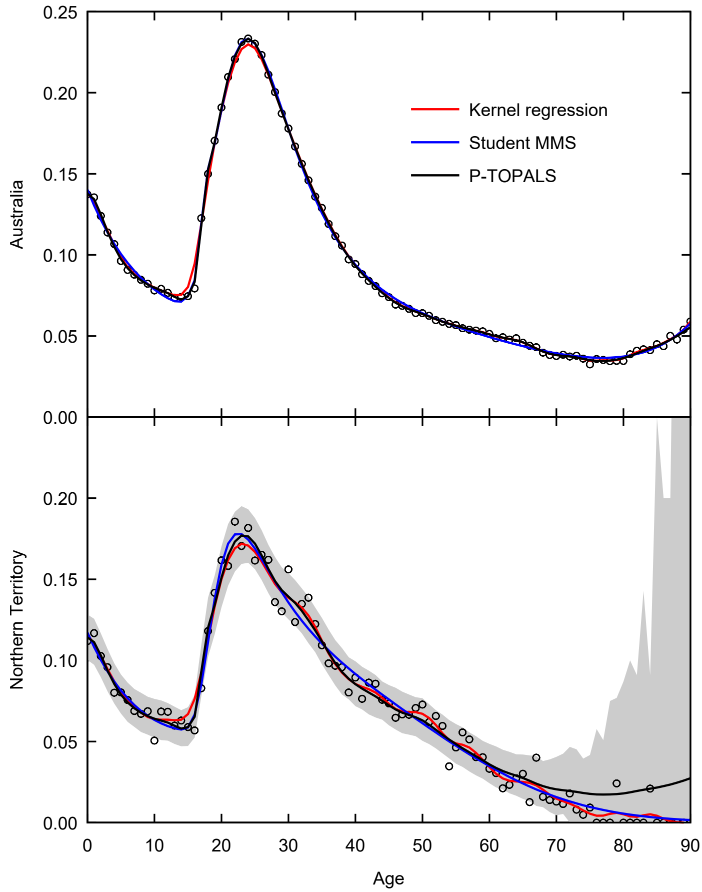
7. Application to intrastate migration

Intrastate migration intensity measures the movement of people between administrative regions within a state. Like interstate migration intensity, it is obtained by dividing the number of people who have made such moves by the relevant population-at-risk, and it can be defined at national, state, or sub-state levels. Data from the 2016 Australian Census of Population and Housing was used to calculate age-specific schedules for intrastate migration between SA2 areas for Australia, the Northern Territory, and two SA2 regions, Nightcliff and Moil, over the one-year interval 2015–2016. Observed and smoothed schedules are shown in Figures 10 and 11 and summary statistics for each fit are given in Table 3. For the calculation of $\bar{P}$, I used the P-TOPALS fit in the top panel of Figure 10.

Comparing the top panel of Figure 10, with interstate intensities shown in the top panel of Figure 2 we see that the intrastate migration curve has a greater overall level, does not have a student peak or as prominent a retirement peak, but does show an increase in intensities after age 80 associated with accommodation-related moves to access aged care or to live with family members (Rogers and Castro 1981; Rogers and Watkins 1987). On average over all ages, the quality of kernel regression and student MMS fits is comparable, with similar values for dev_0. Although there is no student peak, including the student component in Equation (25) does give MMS a better fit than kernel regression over ages 14 to 30, but kernel regression gives a better fit after age 30 and before age 14, where the childhood migration intensities appear to show deviations from the MMS profile.

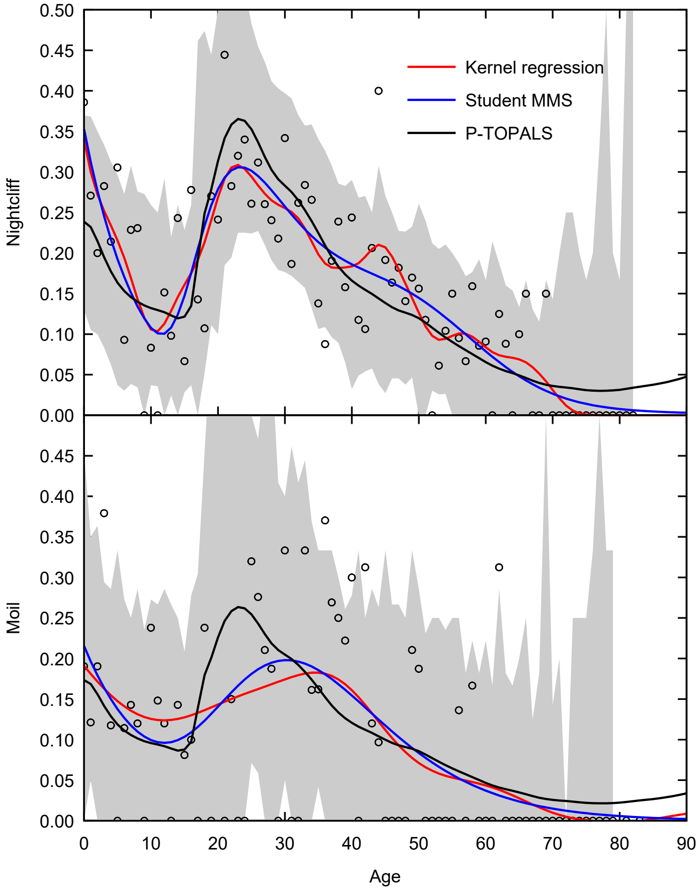
As in Section 4, P-TOPALS can be used to improve an MMS fit. In this case I used the MMS fit as the standard, cubic basis splines with knots spaced two and a half years apart, and a linear penalty with the penalty size determined by the BIC condition. We see from the top panel of Figure 10 that P-TOPALS fits as well as MMS over ages 14–30 and as well as kernel regression over all other ages and has a significantly better global fit as measured by dev_0.

Figure 10: Australian and Northern Territory intrastate migration probabilities between SA2 areas 2015–2016, three smoothing methods



Note: Age is in completed years at the beginning of the migration interval. Grey area, 95% confidence interval for observed intensities based on P-TOPALS fit.
Source: Based on ABS data.

Figure 11: Nightcliff and Moil intrastate out-migration probabilities 2015–2016, three smoothing methods



Note: Age is in completed years at the beginning of the migration interval. Grey area, 95% confidence interval for observed intensities based on P-TOPALS fit.
Source: Based on ABS data.

Table 3: Summary statistics for three smoothing methods applied to intrastate migration for national, state, and sub-state areas, 2015–2016

Area	$\bar{N}$	dev_0			$\bar{P}$			Notes			Fig.
		K	M	P	K	M	P	K	M	P	
Australia	254,444	1,120	1,068	302	0	0	0				10
Northern Territory	2,683	157	197	234	2	2	1				10
Nightcliff	51	96	110	132	4	3	1	X	X		11
Moil	25	201	204	219	7	6	1	X	X		11

Note: Quantity $\bar{N}$ is the average population per single year of age. Goodness-of-fit measure dev_0 is the Poisson deviance for all ages given by Equation (27). Quantity $\bar{P}$ measures the percentage of a schedules' profile that differs from the intrastate profile given in the top panel of Figure 10. K, kernel regression; M, student model migration schedule; P, P-TOPALS. The "Notes" column gives the author's assessment of a fit's deficiencies, if any: S, over-smoothing the student peak; E, under-smoothing advanced ages; X, implausible shape.

In the fits to Northern Territory intrastate migration (Figure 10) and out-migration from Nightcliff and Moil (Figure 11) I have followed the same strategy as in Section 6: kernel regression with linear polynomials, a Gaussian kernel, and a global bandwidth calculated using the rule-of-thumb method; student MMS with starting parameter values taken from the fitted values of the enclosing area; P-TOPALS with standard taken from the enclosing area, quadratic basis splines, and penalty chosen using BIC. We see from Table 3 that for each of these areas, kernel regression gives the best global fit as measured by dev_0, although this is accompanied by the highest value for $\bar{P}$ and, in my opinion, implausible shapes for Nightcliff and Moil. For each of the areas, P-TOPALS has the highest value dev_0 but the profiles are plausible, deviations between observed and smoothed probabilities outside the 95% confidence interval are few, and $\bar{P}$ is consistently low. In terms of dev_0 and $\bar{P}$, student MMS has values between kernel regression and P-TOPALS but also gives implausible shapes for Nightcliff and Moil.

8. Discussion and conclusion

This paper proposes a new method that enables a good estimation of the high-curvature portion of the curve at young adult ages as well as a sensitive modelling of intensities beyond the labour force peak. Using examples of Australian interstate migration, in- and out-migration for its eight states and territories, and four sub-state areas, analysis has shown that P-TOPALS can provide an accurate representation of the migration profile and a robust treatment of sample noise for small populations. Kernel regression and MMS also have their strengths, and there were cases in Sections 4 to 7 where P-TOPALS did not perform better than one of these methods. It should therefore be seen as a useful addition to the applied demographer's toolbox rather than a replacement for these existing methods.

Bernard and Bell (2015) have done a thorough study of the comparative strengths of model migration schedules, cubic splines, and kernel regression for smoothing purposes and the results in this paper are consistent with their findings. Their conclusion that kernel regression and cubic spline are preferable for most countries was based on tests using aggregated five-year migration probabilities. I have also found that kernel regression is more accurate than MMS for five-year interstate probabilities, but for one-year probabilities it could not capture the highly age-concentrated student migration peak.

The main strength of P-TOPALS for generating smooth curves is that it allows users to combine parametric and non-parametric approaches and can be viewed either as a framework for correcting a parametric fit or as means of adding non-polynomial elements to a non-parametric one. Another one of its strengths is its ability to account for the increase in irregularity of observed intensities as population exposed to the risk of moving decreases with age. Ease of use is an important consideration, and P-TOPALS does require more from the user than kernel regression but not as much as student MMS in the sense that users do need to specify a standard curve and an exposure curve but are not faced with the non-trivial problem of adequate starting values for parameters and strategies for guiding them to the best-fit solution.

This paper has focused on graduating transition-type data reported by single year of age. There are a number of paths for further investigation. First, can P-TOPALS be generalised to handle grouped probabilities for countries that report internal mobility using abridged ages? Second, how does the framework need to be extended to handle destination-specific out-migration probabilities of the sort needed for the calculation of multi-regional life tables? Third, when smoothing migration profiles for regions at the sub-state level, what is a good method for choosing the standard? As discussed in Section 3.3, as sample population decreases, the standard is increasingly used to impose an age pattern. Rogers, Little, and Raymer (2010) give three methods for imposing age structure and in their terminology the approach used in Sections 6 and 7 is an example of the regional membership method. Are there cases where family membership or temporal aggregation methods are more appropriate?

9. Acknowledgments

I am grateful to Dr. Tom Wilson for his helpful comments on an earlier draft of this paper. All errors and shortcomings remain my own.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6): 716–723. doi:10.1109/TAC.1974.1100705.

Australian Bureau of Statistics (ABS) (2016). Australian statistical geography standard (ASGS): Volume 1 - Main structure and greater capital city statistical areas. Canberra: ABS (cat. 1270.0.55.001).

Australian Bureau of Statistics (ABS) (2019). TableBuilder, user guide. Canberra: ABS (cat. 1406.0.55.005).

Bell, M., Blake, M., Boyle, P., Duke-Williams, O., Rees, P., Stillwell, J., and Hugo, G. (2002). Cross-national comparison of internal migration: Issues and measures. *Journal of the Royal Statistical Society, Series A* 165(3): 435–464. doi:10.1111/1467-985X.00247.

Bernard, A. and Bell, M. (2015). Smoothing internal migration age profiles for comparative research. *Demographic Research* 32(33): 915–948. doi:10.4054/DemRes.2015.32.33.

Bernard, A., Bell, M., and Charles-Edwards, E. (2014). Life course transitions and the age profile of internal migration. *Population and Development Review* 40(2): 213–239. doi:10.1111/j.1728-4457.2014.00671.x.

Congdon, P. (2008). Models for migration schedules: A Bayesian perspective with applications to flows between Scotland and Wales. In: Raymer, J. and Willekens, F. (eds.). *International migration in Europe: Data, models and estimates*. Chichester: John Wiley and Sons: 193–205. doi:10.1002/9780470985557.ch9.

de Beer, J. (2011). A new relational method for smoothing and projecting age-specific fertility rates: TOPALS. *Demographic Research* 24(18): 409–454. doi:10.4054/DemRes.2011.24.18.

de Beer, J. (2012). Smoothing and projecting age-specific probabilities of death by TOPALS. *Demographic Research* 27(20): 543–592. doi:10.4054/DemRes.2012.27.20.

de Boor, C. (2001). *A practical guide to splines*. New York: Springer. doi:10.2307/2006241.

Eilers, P.H.C. and Marx, B.D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* 11(2): 89–121. doi:10.1214/ss/1038425655.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*. London: Chapman and Hall.

Fraser, B. and Wooton, J. (2005). *A proposed method for confidentialising tabular output* *to protect against differencing*. Paper presented at Joint UNECE/Eurostat work session on statistical data confidentiality, Geneva, Switzerland, November 9–11, 2005.

Gonzaga, M.R. and Schmertmann, C.P. (2016). Estimating age- and sex-specific mortality rates for small areas with TOPALS regression: An application to Brazil in 2010. *Revista Brasileira de Estudos de Populacao* 33(3): 629–652. doi:10.20947/S0102-30982016c0009.

Raymer, J. and Rogers, A. (2008). Applying model migration schedules to represent age-specific migration flows. In: Raymer, J. and Willekens, F. (eds.). *International migration in Europe: Data, models and estimates*. Chichester: John Wiley and Sons: 175–192. doi:10.1002/9780470985557.ch8.

Rees, P., Bell, M., Duke-Williams, O., and Blake, M. (2002). Problems and solutions in the measurement of migration intensities, Australia and Britain compared. *Population Studies* 54(2): 207–222. doi:10.1080/713779082.

Rees, P.H. (1977). The measurement of migration, from census data and other sources. *Environment and Planning A* 9(3): 247–272. doi:10.1068/a090247.

Rogers, A. and Castro, L.J. (1981). Model migration schedules. Laxenburg: International Institute for Applied Systems (Research Report RR–81–30).

Rogers, A., Little, J., and Raymer, J. (2010). *The indirect estimation of migration*. New York: Springer. doi:10.1007/978-90-481-8915-1.

Rogers, A., Raquillet, R., and Castro, L.J. (1978). Model migration schedules and their applications. *Environment and Planning A* 10(5): 475–502. doi:10.1068/a100475.

Rogers, A. and Watkins, J. (1987). General versus elderly interstate migration and population redistribution in the United States. *Research on Aging* 9(4): 483–529. doi:10.1177/0164027587094002.

Ruppert, D., Sheather, S.J., and Wand, M.P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association* 90(432): 257–1270. doi:10.1080/01621459.1995.10476630.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* 6(2): 461–464. doi:10.1214/aos/1176344136.

Thompson, G., Broadfoot, S.J., and Elazar, D.J. (2013). *Methodology for the automatic confidentialisation of statistical outputs from remote servers at the Australian Bureau of Statistics*. Paper presented at Joint UNECE/Eurostat work session on statistical data confidentiality, Ottawa, Canada, October 28–30, 2013.

Wilson, T. (2010). Model migration schedules incorporating student migration peaks. *Demographic Research* 23(8): 191–222. doi:10.4054/DemRes.2010.23.8.

Wooton, J. (2006). Measuring and correcting for information loss in confidentialised census counts. Canberra: Australian Bureau of Statistics (ABS) (cat. 1352.0.55.083).

Appendices

A. Maximising the penalised likelihood function

The maximum of the function Equation (6) satisfies the equation

$$\frac{\partial \mathcal{L}}{\partial \theta} = 0. \tag{A-1}$$

Taking the derivative gives the nonlinear system of equations

$$G'(\theta) \cdot V \cdot ({}_n\tilde{m} - {}_nm) - \lambda D_k' \cdot D_k \cdot \theta = 0, \tag{A-2}$$

where

$$G(\theta) := \frac{1}{{}_nm} \frac{\partial_n m}{\partial \theta}. \tag{A-3}$$

Let $G_x(\theta)$ denote the $(x+1)$th row vector of $G(\theta)$ and B_j the $(j+1)$th row vector of B. Taking the derivatives of Equation (4) and (5) gives

$$G_x(\theta) = \frac{1 - {}_nm_x}{{}_nm_x} \left(\sum_{x \le j < x+n} \frac{m_j}{1 - m_j} B_j \right). \tag{A-4}$$

To solve Equation (A-2) I use approximations

$$\begin{aligned} {}_nm(\bar{\theta}) &\approx {}_nm + {}_nm\, G \cdot (\bar{\theta} - \theta), && \text{(A-5)} \\ G(\bar{\theta}) &\approx G && \text{(A-6)} \end{aligned}$$

which when substituted into Equation (A-2) give the linear iteration Equation (8).

B. Smoothed migration curves by state

This section summarises the methods I used to obtain the smoothed one-year migration curves shown in Figures 5, and B-1 to B-7. For the kernel regression fits I used linear polynomials and a Gaussian kernel with a global bandwidth calculated using Ruppert, Sheather, and Wand's (1995) rule-of-thumb method.

For most of the student MMS fits I followed the same procedure used in Section 4 for interstate migration. This procedure did not converge to a sensible solution for Australian Capital Territory in-migration (Figure B-6), although it did for 2006 and 2011 data. In this case I found it necessary to first fit the model to in-migration aggregated over 2006, 2011, and 2016 censuses, and then use the fitted parameters as a starting point for a fit of all parameters to the 2016 census data. For Western Australia, South Australia, and Queensland out-migration (Figures B-3 to B-5) all parameters were fitted except the position of the student peak, which was held fixed at $\mu_5 = 17.5$. In-migration intensities for Western Australia and Northern Territory did not exhibit a strong student peak, and in these two cases I used the standard MMS (Rogers and Watkins 1987)

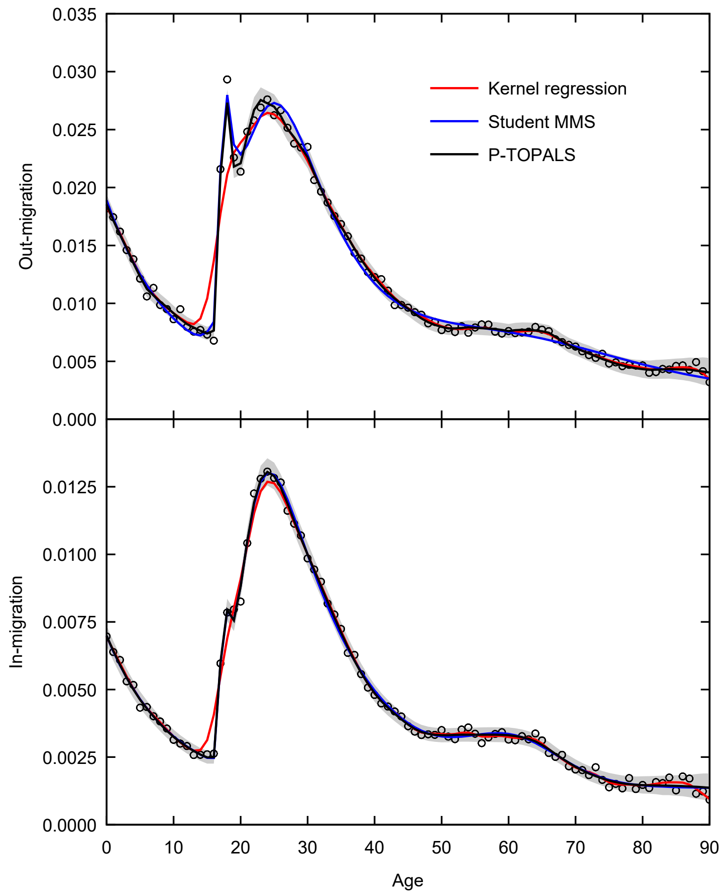
$$
\begin{aligned}
m_x \quad = \quad & a_1 \exp(-\alpha_1 x) && \textit{(childhood)} \\
& + a_2 \exp\left(-\alpha_2(x-\mu_2) - e^{-\lambda_2(x-\mu_2)}\right) && \textit{(labor force)} \\
& + a_3 \exp\left(-\alpha_3(x-\mu_3) - e^{-\lambda_3(x-\mu_3)}\right) && \textit{(retirement)} \\
& + a_4 \exp\left(\alpha_4 x\right) && \textit{(elderly)} \\
& + c && \textit{(constant)},
\end{aligned}
\qquad \text{(B-1)}
$$

which differs from the student MMS in not having a student component and having a non-symmetrical functional form for the retirement component.

For the P-TOPALS fits, I used four types of standards. Since interstate migration is an average of both in- and out-migration, it was a natural choice as the standard for some cases (see out-migration in Figures B-1, B-2, B-6, and B-7). For cases where the student peak of the interstate curve was either very large or very small compared to the state, I used the student MMS fit as the standard if it was a reasonable fit to the student peak (see out-migration in Figures B-3 to B-5 and in-migration, in Figures B-1 to B-3). When neither the interstate curve nor the student MMS curve gave good fits to the student peak, I created a standard curve using time-aggregated migration data, from 2006, 2011, and 2016 censuses, first fit with student MMS and then corrected using P-TOPALS as in Section 4 (see in-migration in Figures 5 and B-6). For Tasmania in-migration this approach led P-TOPALS to have a worse fit than MMS for 2016 data but more realistic fits for 2006 and 2011 data which has a less prominent student peak. For Western Australia and Northern Territory in-migration, which did not exhibit a student peak or excessive levels of sample noise, I used a flat standard $\hat{m} = 1$.

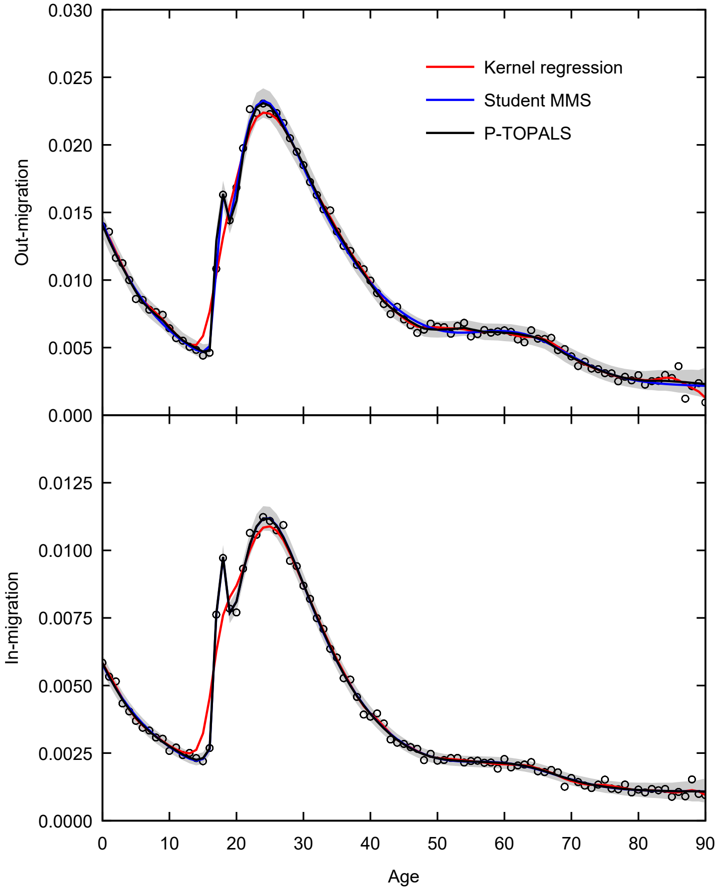
BIC was usually a good choice for the penalty selection criterion. Sometimes it appeared to over-smooth, in which case I chose AIC (see out-migration in Figures 5, B-4, and B-5 and in-migration in Figures B-1 to B-3). Linear B-splines were usually adequate but in some cases their piece-wise linear form led to kinks at the spline knots, and for these I used quadratic splines (see out-migration in Figures 5, B-4, and B-5 and in-migration in Figures B-1 to B-4 and B-7).

Figure B-1: New South Wales one-year migration probabilities 2015–2016 by age, three smoothing methods



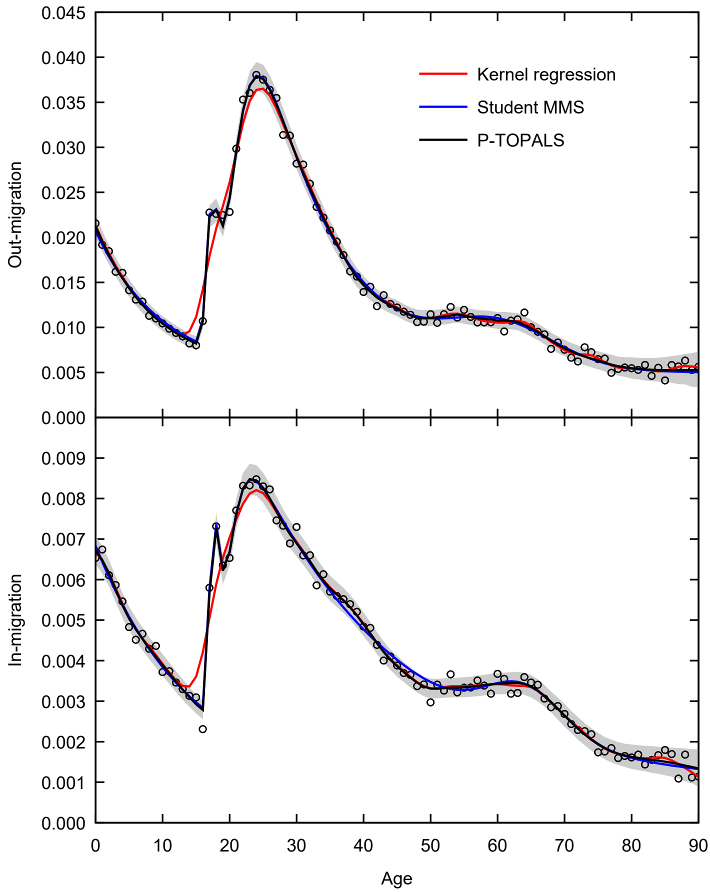
Note: Age is in completed years at the beginning of the migration interval. Grey area, 95% confidence interval for observed intensities based on P-TOPALS fit.
Source: Based on ABS data.

Figure B-2: Victoria one-year migration probabilities 2015–2016 by age, three smoothing methods



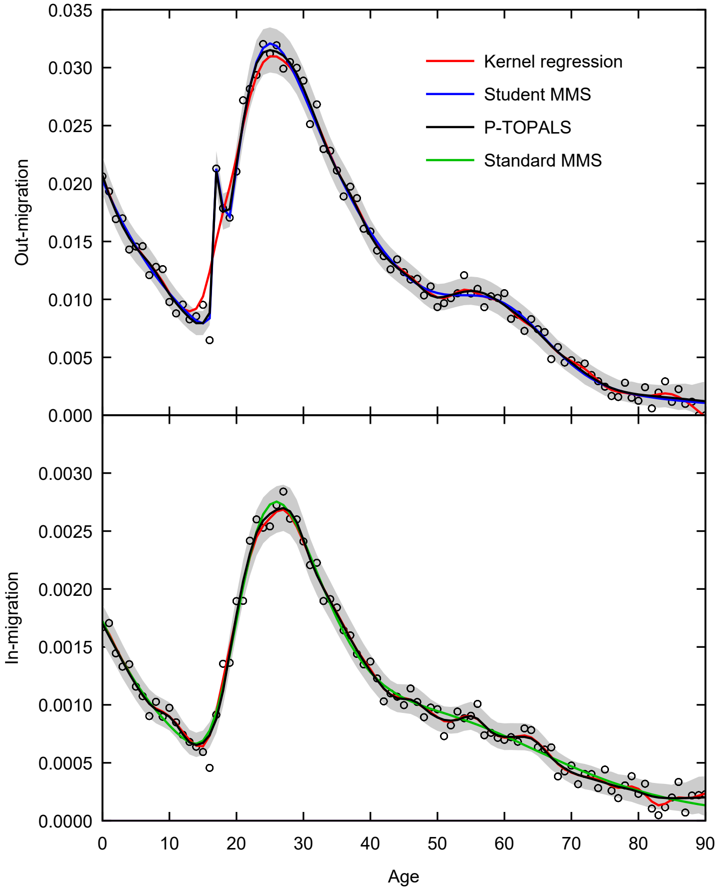
Note: Age is in completed years at the beginning of the migration interval. Grey area, 95% confidence interval for observed intensities based on P-TOPALS fit.
Source: Based on ABS data.

Figure B-3: Queensland one-year migration probabilities 2015–2016 by age, three smoothing methods



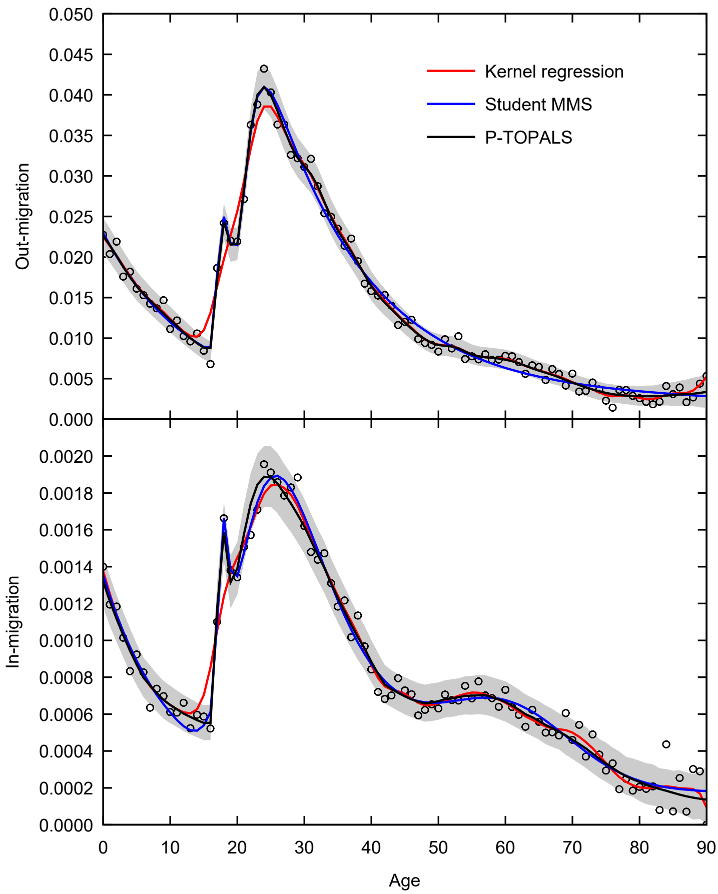
Note: Age is in completed years at the beginning of the migration interval. Grey area, 95% confidence interval for observed intensities based on P-TOPALS fit.
Source: Based on ABS data.

Figure B-4: Western Australia one-year migration probabilities 2015–2016 by age, three smoothing methods



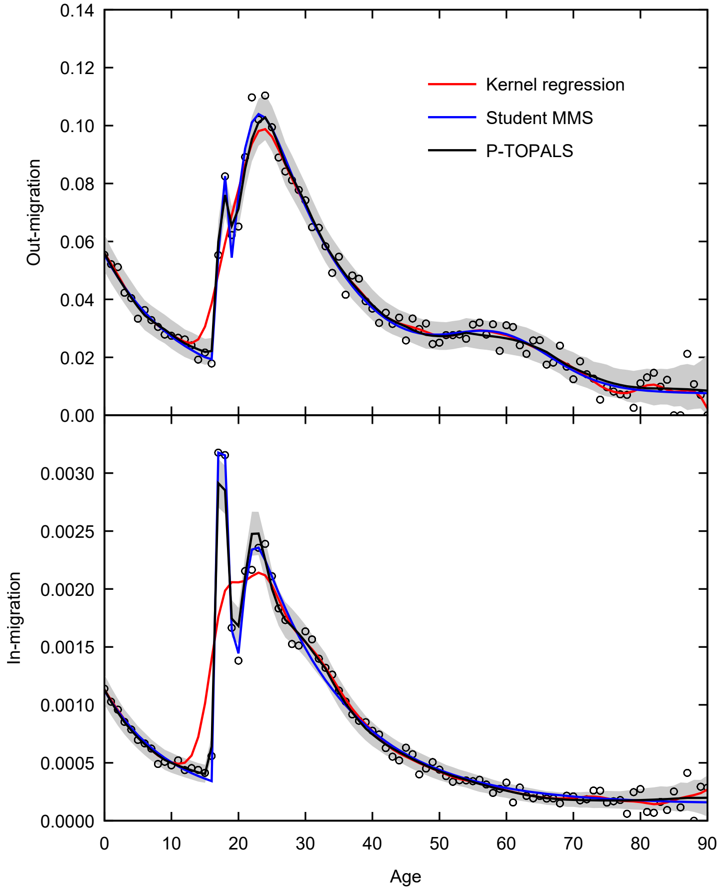
Note: Age is in completed years at the beginning of the migration interval. Grey area, 95% confidence interval for observed intensities based on P-TOPALS fit.
Source: Based on ABS data.

Figure B-5: South Australia one-year migration probabilities 2015–2016 by age, three smoothing methods



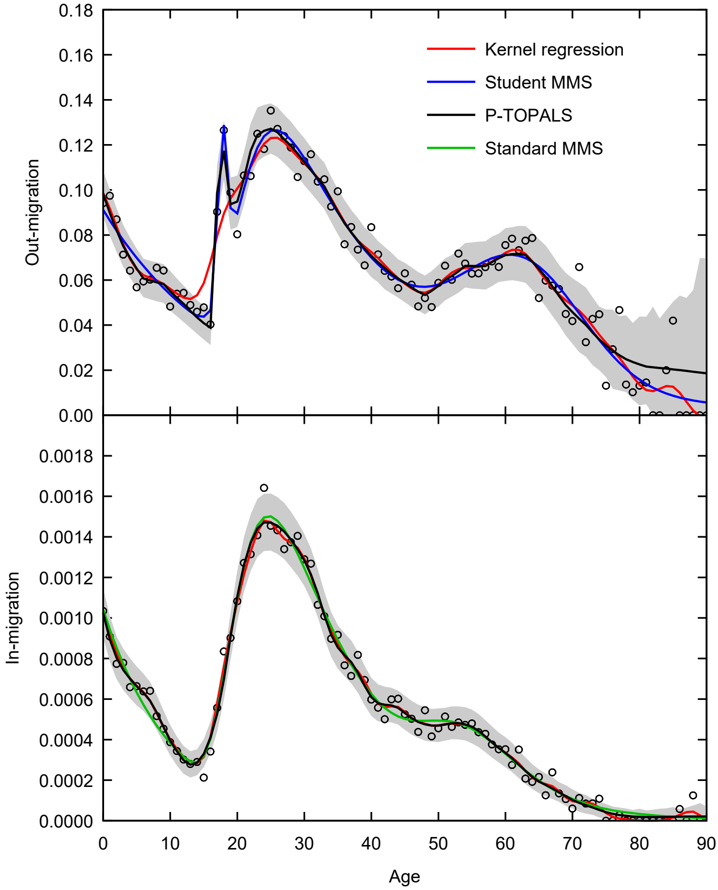
Note: Age is in completed years at the beginning of the migration interval. Grey area, 95% confidence interval for observed intensities based on P-TOPALS fit.
Source: Based on ABS data.

Figure B-6: The Australian Capital Territory one-year migration probabilities 2015–2016 by age, three smoothing methods



Note: Age is in completed years at the beginning of the migration interval. Grey area, 95% confidence interval for observed intensities based on P-TOPALS fit.
Source: Based on ABS data.

Figure B-7: Northern Territory one-year migration probabilities 2015–2016 by age, three smoothing methods



Note: Age is in completed years at the beginning of the migration interval. Grey area, 95% confidence interval for observed intensities based on P-TOPALS fit.
Source: Based on ABS data.