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.

DATA AND 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 which can be solved using linear techniques. I compare it to smoothing by model migration schedules and kernel regression using 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.

CONCLUSION
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 aggregate 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 both parametric and non-parametric approaches to give a good general-purpose smoother.

[1]Northern Institute, Charles Darwin University, Australia. E-mail: sigurd.dyrting@cdu.edu.au

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 in- and out-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). In the case of graduation with parametric functions it can be used to reduce the effective dimension of a migration curve as a precursor to calibrating a projection model (Girosi and King 2008). 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 twenties 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 applied to internal and international migration flows in developed as well as less-developed countries 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). It is well known that 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) has investigated Bayesian approaches to migration graduation and concluded that non-parametric models could detect features in the migration data that MMS could not.

A good general-purpose smoothing method for migration probabilities must serve more than one master. It needs to be accurate when there is little noise and realistic when there is a lot. It needs to work well for all ages for which there is data and when there is no data extrapolate in a manner that is easy to control. It must work equally well for single-year and multi-year data and allow a like-for-like comparison between the two. We shall see that neither MMS nor kernel regression alone are satisfactory in this respect. 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 such as 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 for migration graduation that brings together the strengths of both parametric and non-parametric approaches to give a good general-purpose smoother. My approach is to combine de Beer (2011)'s relational Tool For Population Analysis using Linear Splines (TOPALS) with Eilers and Marx (1996)'s 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 census-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 aggregate 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 performance under conditions of increasing levels of noise and flexibility in fitting a range of profile shapes.

2 Smoothing Migration Probabilities

Internal migration data collected by national census 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. Countries typically collect data for fixed intervals of one ($n = 1$) or five ($n = 5$) years (Bell et al. 2014). 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

$$ {}_n m = \begin{bmatrix} {}_n m_0 \\ \vdots \\ {}_n m_\omega \end{bmatrix} \tag{3} $$

that in some sense fits ${}_n\tilde{m}$ and is smooth. Conceptually we regard the vector ${}_n\tilde{m}$ as consisting of persistent components ${}_n m$ 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 relational and spline methods are in a sense complementary and that by expressing age-specific rates as a product of a standard and a spline one creates a framework greater than the sum of its parts: 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 equal 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 ${}_n m$ in terms of probabilities at one-year intervals m_k

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

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 m 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

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

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 is a penalty that as λ becomes larger forces the difference term $B \cdot \theta$ to tend to a polynomial of degree $k - 1$.

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

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

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

and

$$W(\theta) = \text{diag}({}_nm\,N), \tag{11}$$

$$V = \text{diag}(N). \tag{12}$$

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

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

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 (1978)'s Bayesian information criterion (BIC): λ is found by minimising the function

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

where the first term is the deviance of the fit and

$$\dim(\theta, \lambda) = \mathrm{tr}(H) \tag{15}$$

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

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

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 (1974)'s information criterion (AIC): λ is found by minimising the function

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

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 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 fifteen or fifty 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 then 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 which 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 aggregate interstate migration

Aggregate 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. Because it counts all movers, irrespective of source or destination, it is often used for cross-national comparisons of internal migration (Bell et al. 2002). It is also 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 2016 Census of Population and Housing and used to calculate raw aggregate 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 (1995)'s rule-of-thumb plug-in bandwidth selector. For the MMS fit I chose Wilson (2010)'s sixteen parameter student model

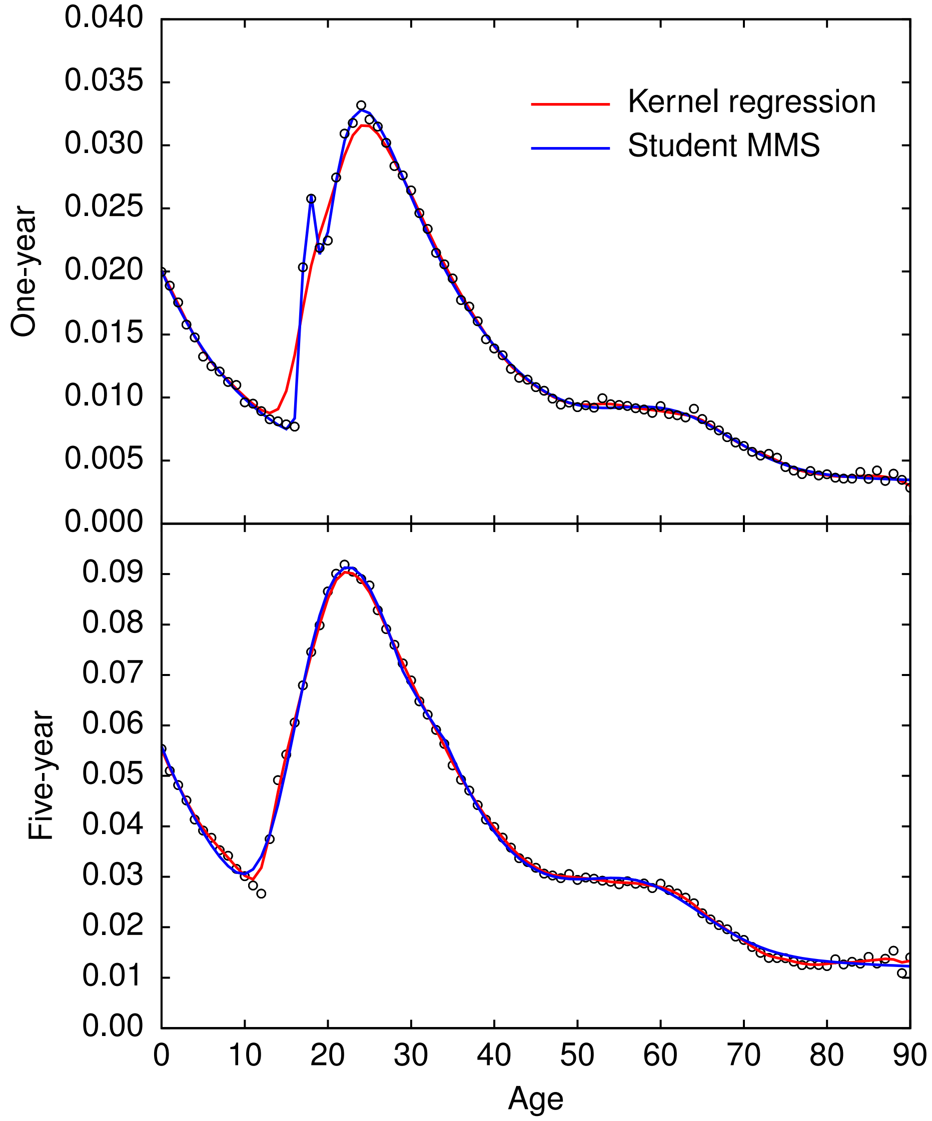
$$
\begin{aligned}
m_x \;=\; & a_1 \exp(-\alpha_1 x) && \textit{(childhood)} \\
& + a_2 \exp\left(-\alpha_2(x-\mu_2) - e^{-\lambda_2(x-\mu_2)}\right) && \textit{(labour force)} \\
& + a_3 \exp\left(-\left(\frac{(x-\mu_3)}{\sigma_3}\right)^2\right) && \textit{(retirement)} \\
& + a_4 \exp(\alpha_4 x)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

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 fourteen 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.

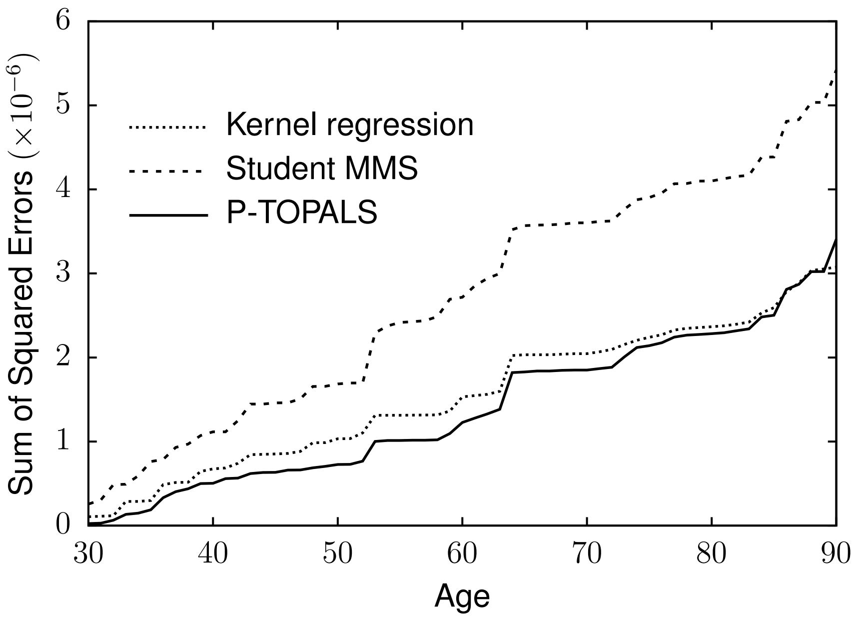
What is perhaps not so clear is that after age 30 kernel regression provides a better fit than MMS. This can be seen in Figure 2 which plots the cumulative sum of squared errors from age 30. The relatively poor performance of MMS over this age range is probably due to limitations imposed by a parametric profile. Thus we see that for one-year aggregate 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

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



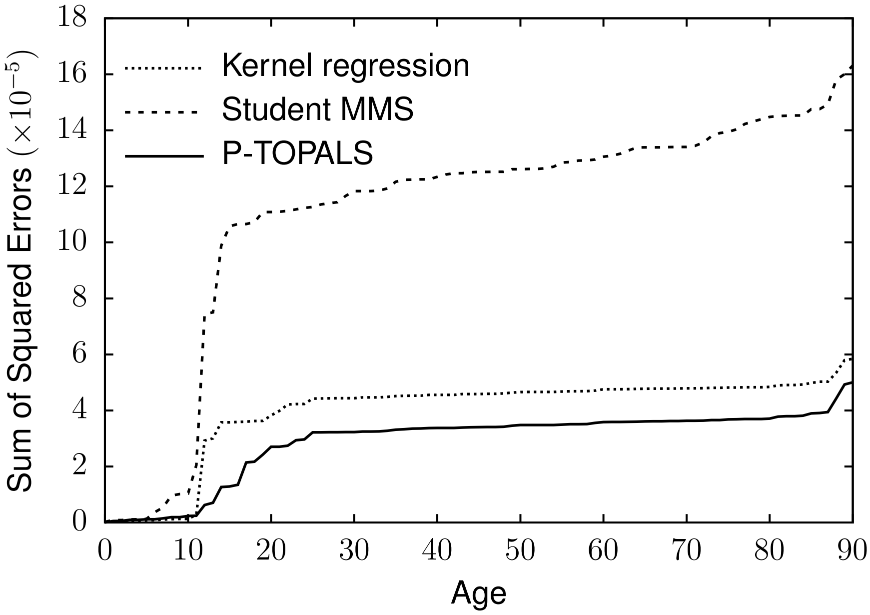
Note: Age is in completed years at the beginning of the migration interval.
Source: Based on Australian Bureau of Statistics data.

Figure 2: Australian aggregate interstate migration 2015-2016, cumulative sum of squared smoothing errors from age 30.



Note: Age is in completed years at the beginning of the migration interval.

Figure 3: Australian aggregate interstate migration 2011-2016, cumulative sum of squared smoothing errors.



Note: Age is in completed years at the beginning of the migration interval.

$\hat{m}$ 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 4 shows the P-TOPALS fit. The fit to student peak has been preserved and as Figure 2 shows the fit after age 30 is now as good as kernel regression.

The bottom panel of Figure 1 shows fits to five-year data. Five-year migration probabilities are preferred for cross-national comparisons because they include the effects of return and onwards migration (Courgeau 1973; Rees 1977) and are therefore a more accurate reflection of spatial changes over the medium term. Comparing one and five-year data we see that the five-year age profile is more smooth. 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. Indeed, a plot of the cumulative sum of squared errors from age 0 given in Figure 3 shows that kernel regression out-performs Student MMS across the entire age range.

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 3 by the sudden increase in the cumulative sum of squared errors 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 4 which has a jump in migration intensity at age seventeen. 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 3 shows there is a more gradual increase in the sum of squared errors over ages ten to twenty and a better performance than both Student MMS and kernel regression across the entire age range.

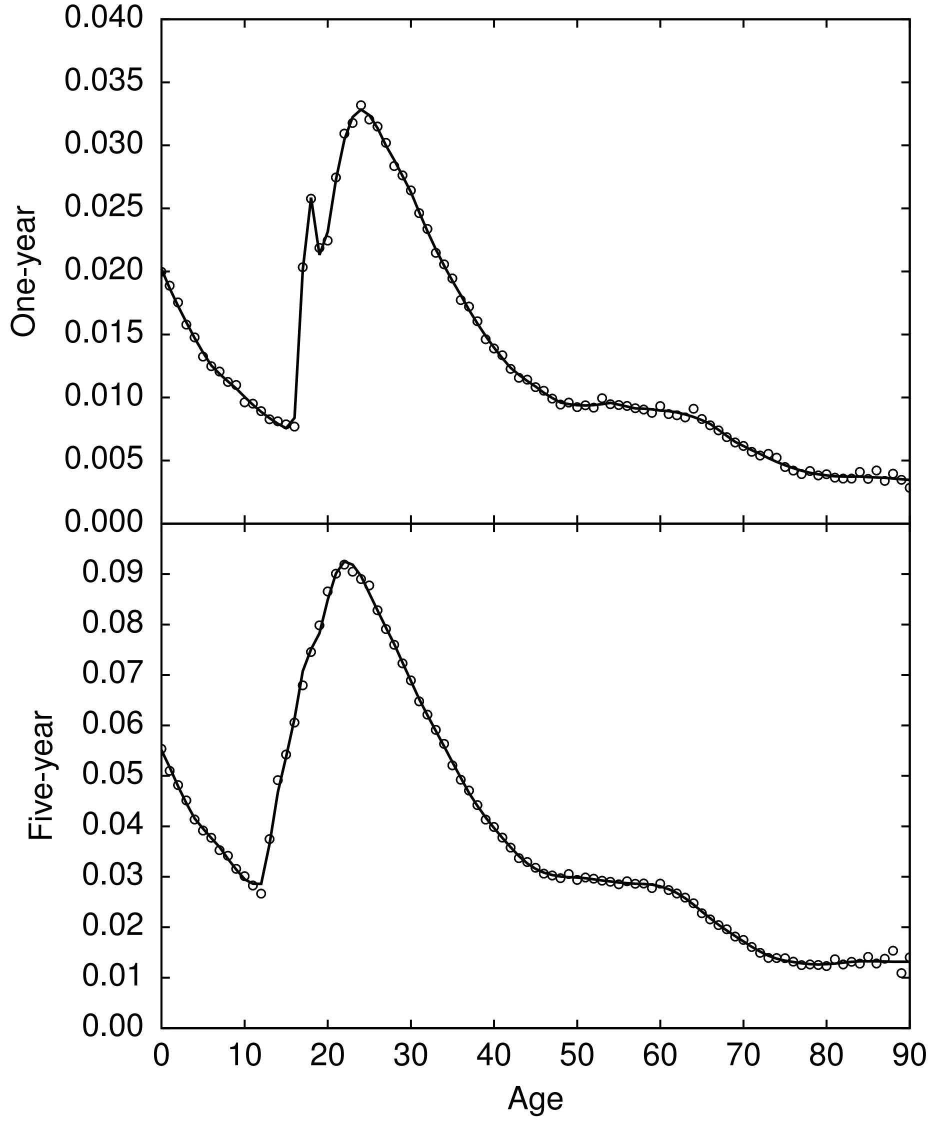
5 Application to individual states

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 without and within the state respectively. Because they count movers to and from a given region they are an important input for subnational population projections. Methods for generating smooth age profiles at the state level must be both flexible and robust because differences in population, economy, climate, and infrastructure affect the occurrence, strength and timing of migration events and lead to significant variations in migration profiles, both by region and by direction.

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 the Northern Territory are given in Appendix B. Since one state's departure is another's arrival it follows that aggregate interstate migration is a weighted average of in-migration or of out-migration, where the weights are the population without (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 this from average.

Appendix B summarises the strategies used to fit the observed probabilities with our three methods. 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 and B-3 and in-migration

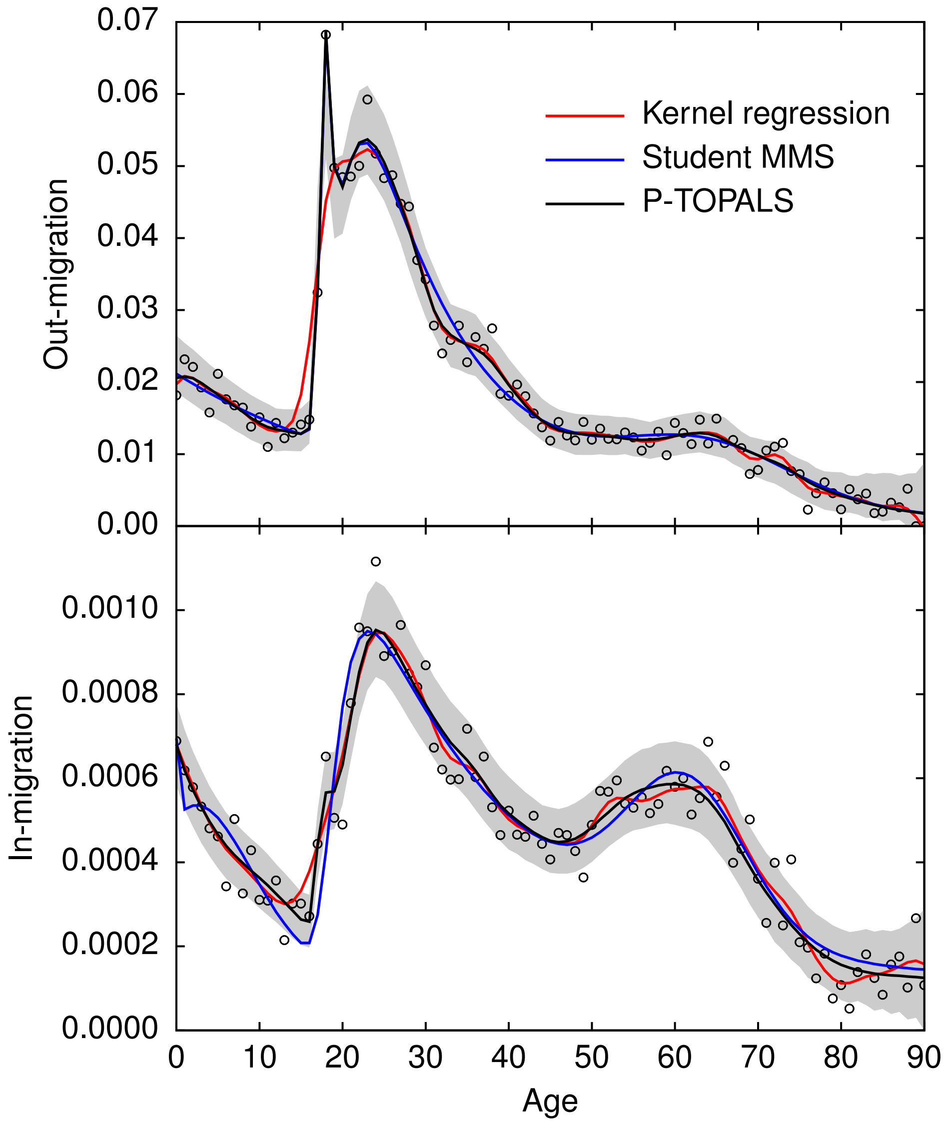
Figure 4: Australian aggregate interstate migration probabilities 2016, smoothing with P-TOPALS.



Note: Age is in completed years at the beginning of the migration interval.

Source: Based on Australian Bureau of Statistics data.

Figure 5: Tasmania 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.

Source: Based on Australian Bureau of Statistics data.

in Figures B-2 and B-4). Kernel regression gives a bad fit to the student peak whenever it is present (see out-migration in Figures 5 and B-1 to B-7 and in-migration in Figures B-2, B-3, B-5 and B-6) but in general provides good fits after it. P-TOPALS gives good fits across all ages, showing remarkably similar age profiles to kernel regression after the labour force peak despite the two methods being based on different algorithms (see out-migration in Figures 5, B-3 and B-7 and in-migration in Figures B-2, B-4 and B-5).

There is an increase in the level of irregularity in observed migration rates as population N decreases from the larger states (see Figures B-5 to B-7) to the middle (see Figures B-3 and B-4) and the smaller ones (see Figures 5, B-1 and B-2). The three graduation methods respond to this in different ways. Student MMS is usually robust but occasionally finds it difficult to distinguish between noise and the student peak (see in-migration in Figures 5, B-1, B-3 and B-4). Kernel regression can have problems smoothing for advances 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 sixty (see out-migration in Figures 5 and B-1 and in-migration in Figures 5, B-1 and B-4). We see that P-TOPALS shows neither the problems of Student MMS in locating the student peak nor problems of kernel regression in giving sensible profiles for advanced ages. The superior performance of P-TOPALS is in part due to the stabilising effect of the standard which, as discussed in Section 3.3, determines the profile of the student peak and, when N is very small, the entire age profile. It is also due to the flexibility of the P-spline framework which allows us to easily incorporate a Poisson model of sample noise so that differences between fitted and observed probabilities are weighted less for ages where the population is low than where it is high.

6 Discussion and conclusion

For countries such as Australia population structure and dynamics at the sub-national level are mainly determined by internal migration and capturing the age-dependent nature of this process through graduation is vital for a range of demographic analyses such as population projection and comparative research. 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 aggregate interstate migration and in- and out-migration for its eight states and territories analysis has shown that P-TOPALS can provide a more accurate representation of the migration profile than both Student MMS and kernel regression and a more robust treatment of sample noise for small populations.

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. They rightly emphasise that the quality of an MMS fit can be very sensitive to the initial values of the parameters which I have also found in some of my fits to state-level data. Their conclusion that kernel regression and cubic spline are preferable for most countries was based on tests using five-year migration probabilities. I have also found that kernel regression is superior to MMS in this case, 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 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 always 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 choosing good 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?

7 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.

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.

Bell, M., Charles-Edwards, E., Kupiszewska, D., Kupiszewski, M., Stillwell, J., and Zhu, Y. (2014). Internal migration data around the world: Assessing contemporary practice. *Population, Space and Place* 21(1): 1–17.

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

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.

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 & Sons: 193–205.

Courgeau, D. (1973). Migrants and migrations. *Population* 28: 95–128.

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

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

de Boor, C. (2001). *A Practical Guide To Splines*. New York: Springer.

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

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. London: Chapman & Hall.

Girosi, F. and King, G. (2008). *Demographic Forecasting*. Princeton: Princeton University Press.

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.

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 & Sons: 175–192.

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* (2): 207–222.

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

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.

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

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.

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.

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

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

A Maximising the penalised likelihood function

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

$$\frac{\partial \mathcal{L}}{\partial \theta} = 0. \tag{27}$$

Taking the derivative gives the nonlinear system of equations

$$G'(\theta) \cdot V \cdot ({}_n\tilde{m} - {}_nm) - \lambda D_k' \cdot D_k \cdot \theta = 0 \tag{28}$$

where

$$G(\theta) := \frac{1}{{}_nm} \frac{\partial {}_nm}{\partial \theta}. \tag{29}$$

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

$$G_x(\theta) = \frac{1 - {}_nm_x}{{}_nm_x} \left(\sum_{x \le j < x+n} \frac{m_j}{1 - m_j} B_j \right). \tag{30}$$

To solve Equation (28) I use approximations

$$
\begin{aligned}
{}_nm(\bar{\theta}) &\approx {}_nm + {}_nm\, G \cdot (\bar{\theta} - \theta) && (31)\\
G(\bar{\theta}) &\approx G && (32)
\end{aligned}
$$

which when substituted into Equation (28) give the linear iteration Equation (8).

B Smoothed migration curves by state

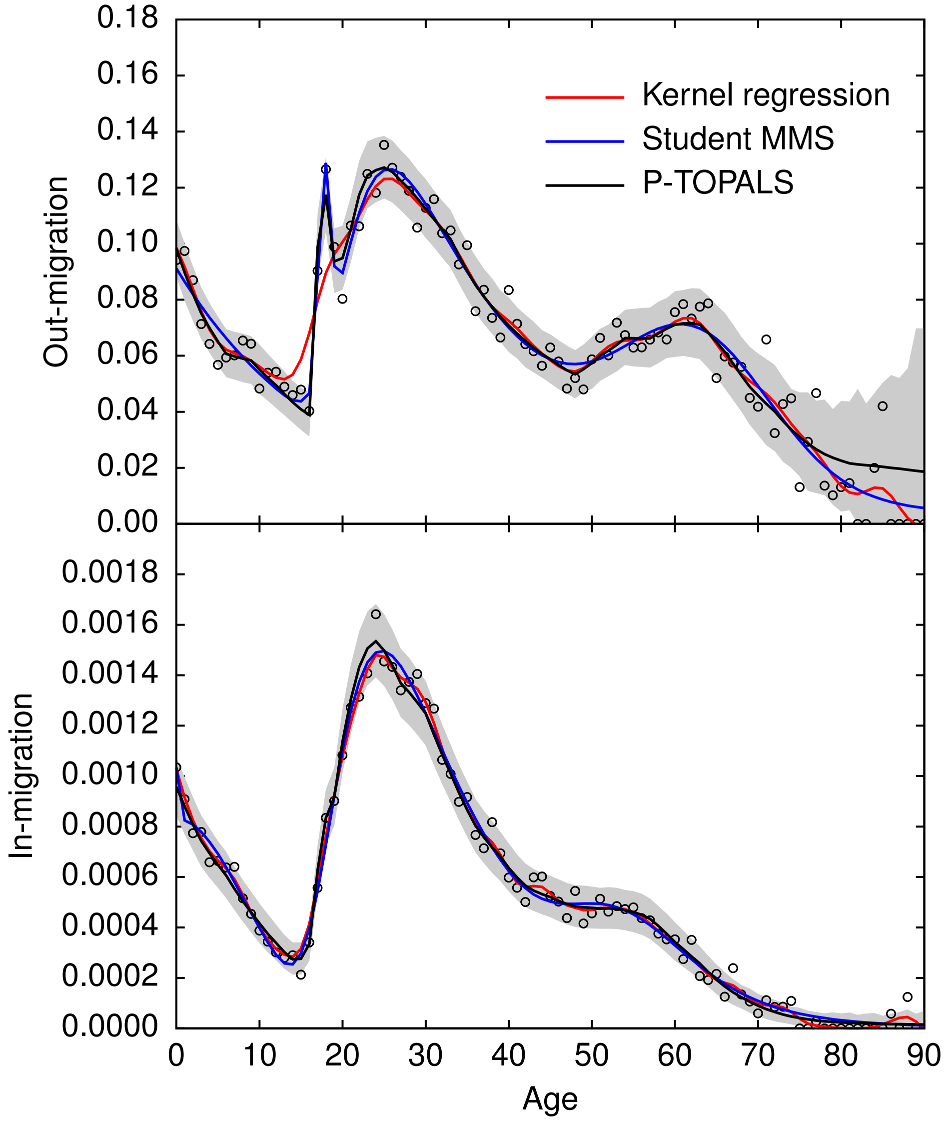
This section summarises the methods I used to obtain the smoothed migration curves shown in Figures 5 and B-1 to B-7. For the kernel regression fits I used linear polynomials and a Gaussian kernel with a global bandwidth calculated using Ruppert, Sheather, and Wand (1995)'s rule-of-thumb method.

For most of the Student MMS fits I followed the same procedure used in Section 4 for aggregate interstate migration. For the Australian Capital Territory in-migration (Figure B-2) I first fitted the model to in-migration aggregated over 2006, 2011, and 2016 censuses and then used the 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 student peak which was held fixed at $\mu_5 = 17.5$.

For the P-TOPALS fits I used three types of standards. Since aggregate 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 aggregate 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 5 and B-3 to B-5 and in-migration in Figures B-5 to B-7). When neither the aggregate 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, B-1 and B-2).

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-3 and B-4 and in-migration in Figures B-5 to B-7). Linear B-splines were usually adequate but in some cases their piece-wise linear form lead to kinks at the spline knots and for these I used quadratic splines (see out-migration in Figures 5, B-3 and B-4 and in-migration in Figures B-5 to B-7).

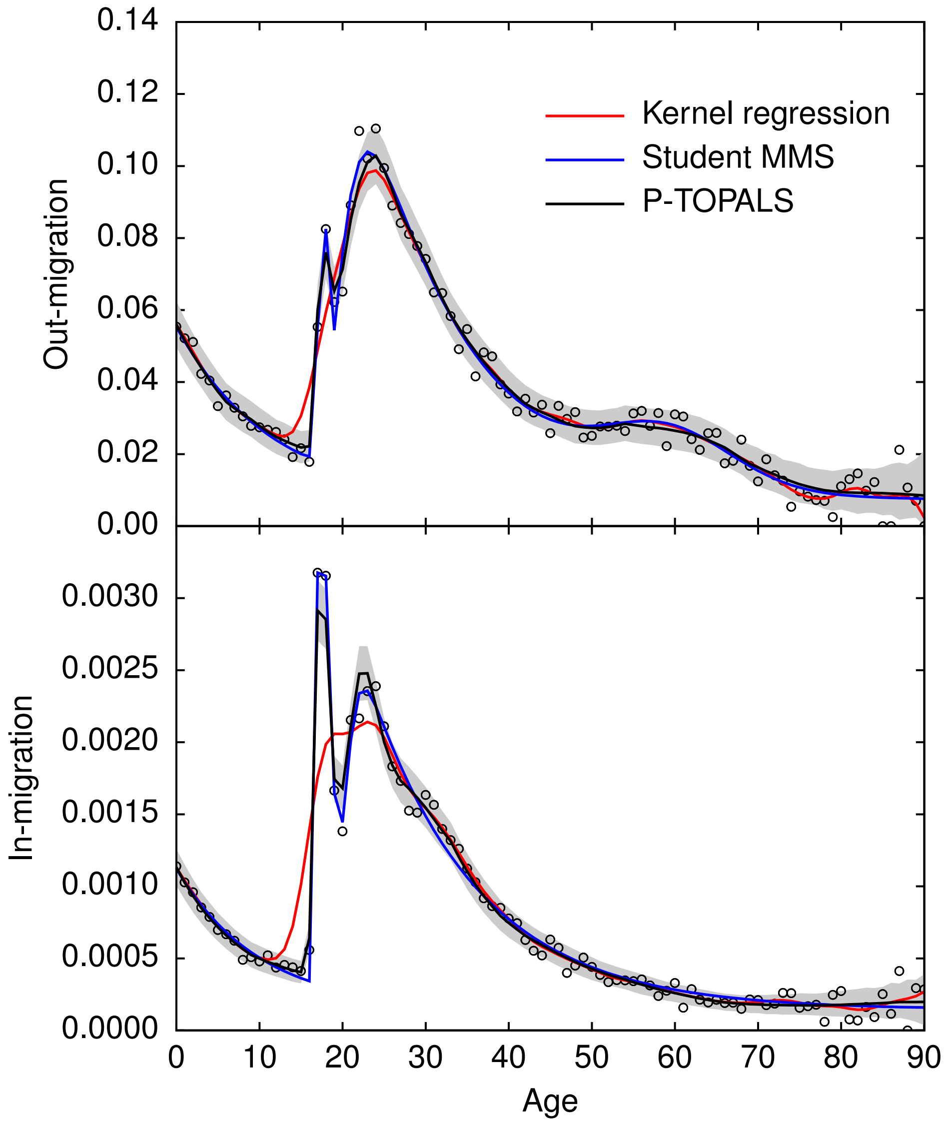
Figure B-1: Northern Territory migration probabilities 2015-2016 by age, three smoothing methods.



Note: Age is in completed years at the beginning of the migration interval. Grey area, 95% confidence interval.

Source: Based on Australian Bureau of Statistics data.

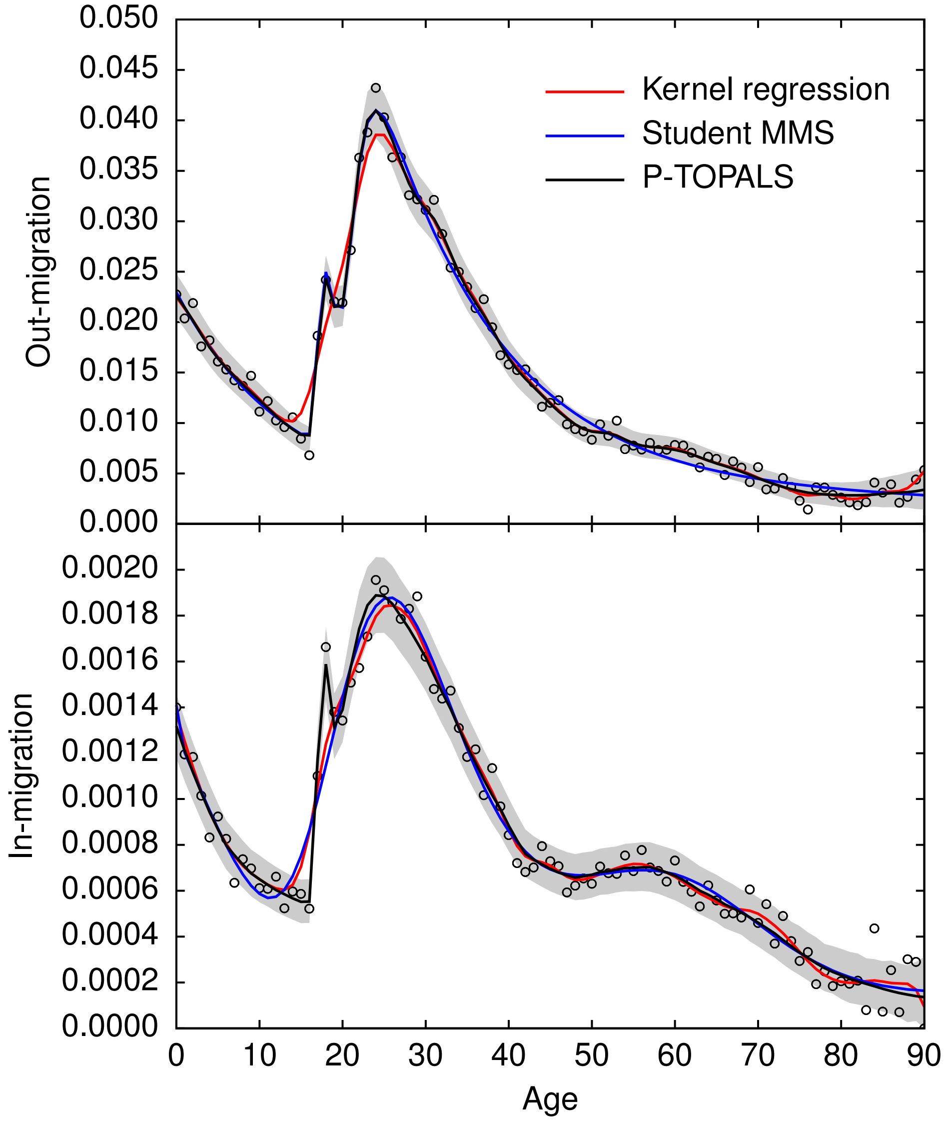
Figure B-2: The Australian Capital Territory migration probabilities 2015-2016 by age, three smoothing methods.



Note: Age is in completed years at the beginning of the migration interval. Grey area, 95% confidence interval.

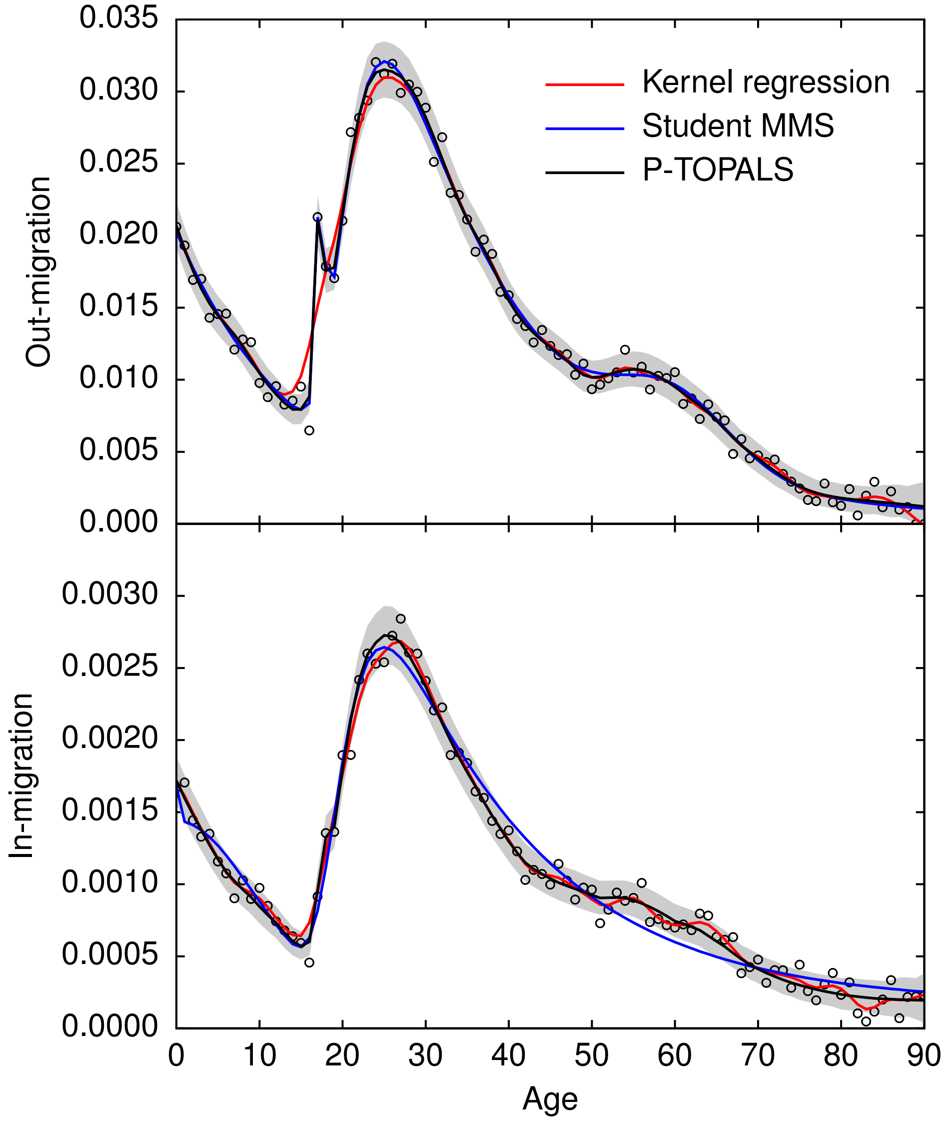
Source: Based on Australian Bureau of Statistics data.

Figure B-3: South Australia 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.
Source: Based on Australian Bureau of Statistics data.

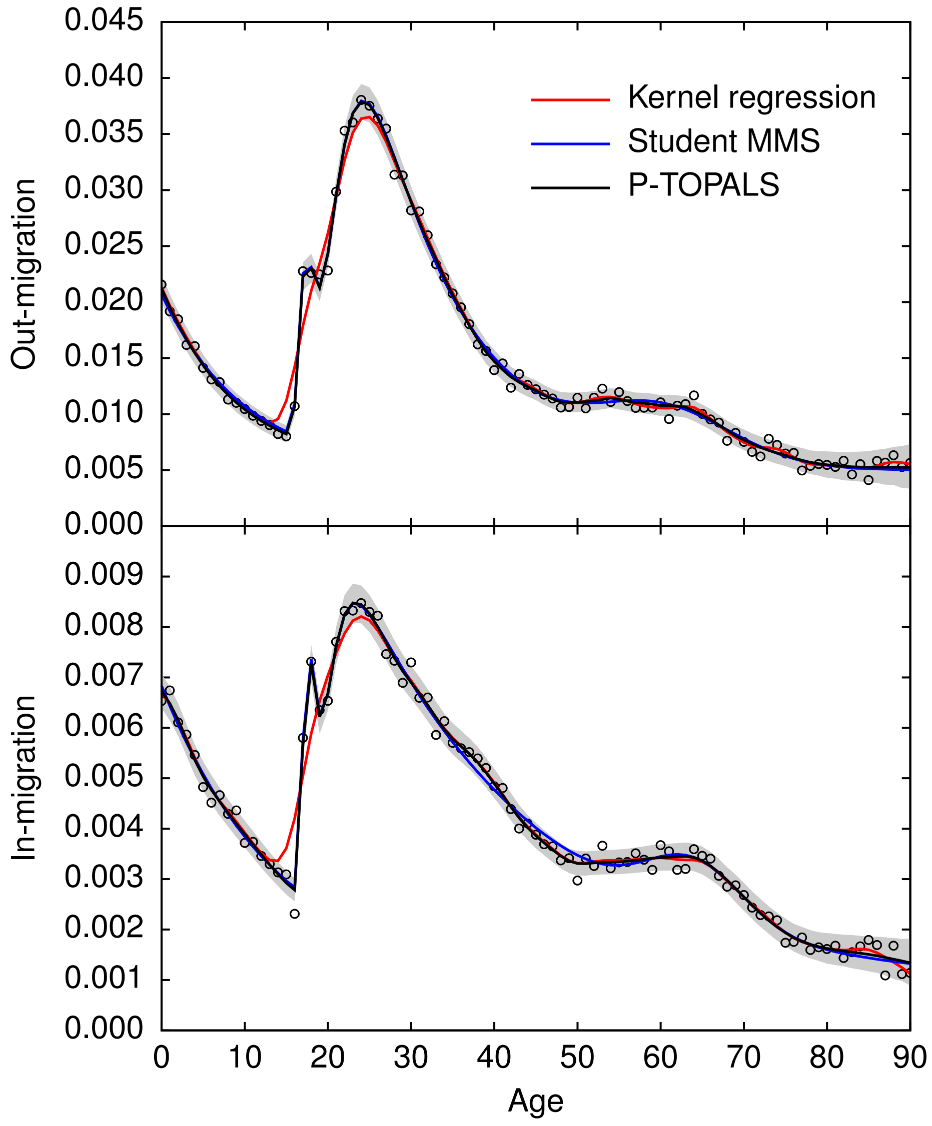
Figure B-4: Western Australia migration probabilities 2015-2016 by age, three smoothing methods.



Note: Age is in completed years at the beginning of the migration interval. Grey area, 95% confidence interval.

Source: Based on Australian Bureau of Statistics data.

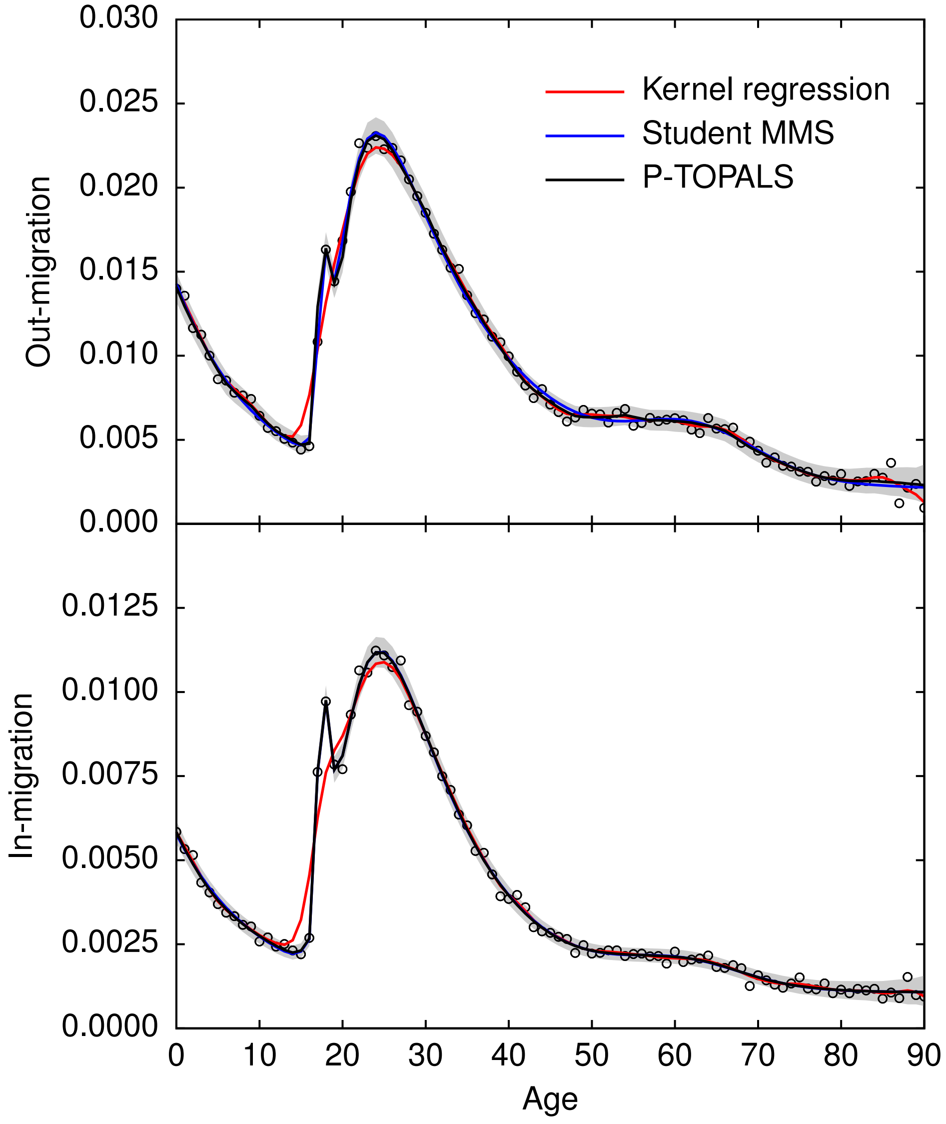
Figure B-5: Queensland migration probabilities 2015-2016 by age, three smoothing methods.



Note: Age is in completed years at the beginning of the migration interval. Grey area, 95% confidence interval.

Source: Based on Australian Bureau of Statistics data.

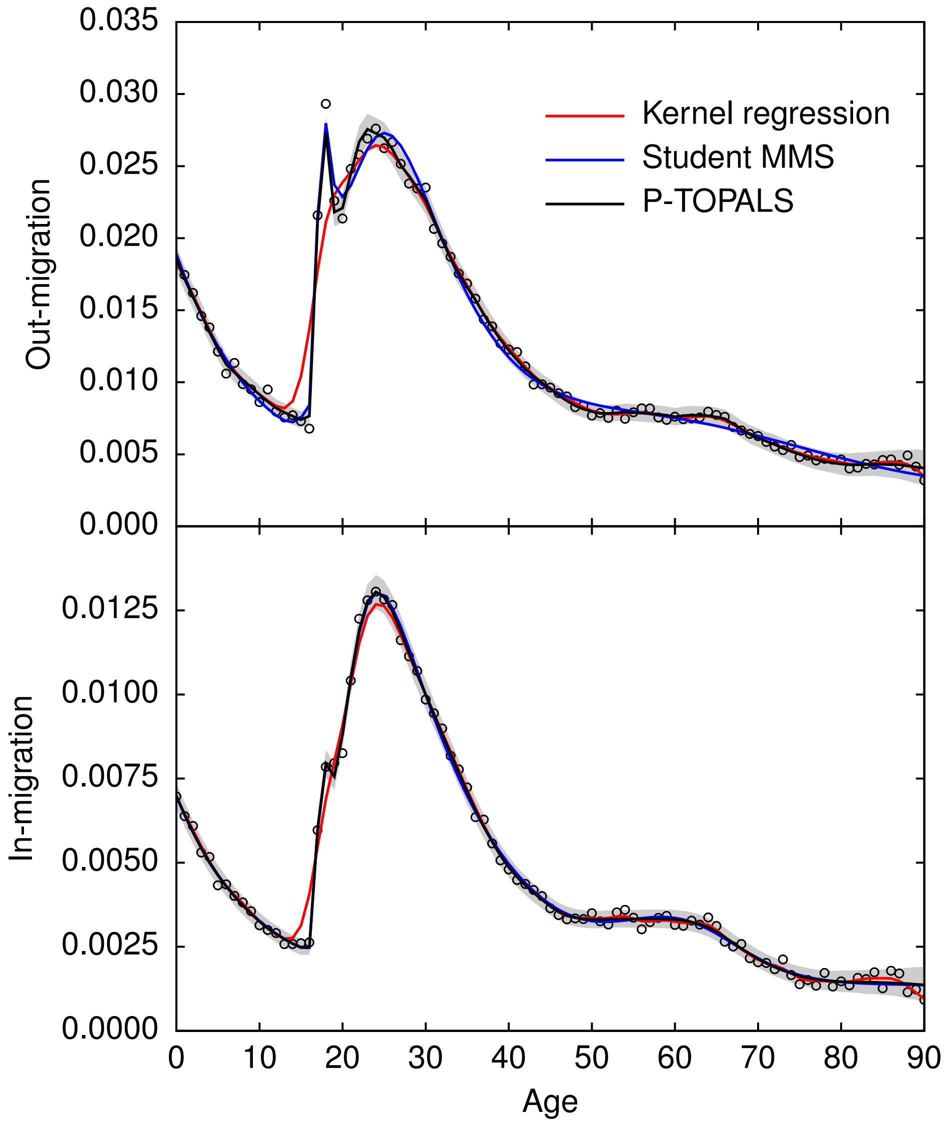
Figure B-6: Victoria migration probabilities 2015-2016 by age, three smoothing methods.



Note: Age is in completed years at the beginning of the migration interval. Grey area, 95% confidence interval.

Source: Based on Australian Bureau of Statistics data.

Figure B-7: New South Wales 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.

Source: Based on Australian Bureau of Statistics data.