



A methodology for projecting sparse populations and its application to remote Indigenous communities

Alfred Michael Dockery¹ · Mark N. Harris² · Nicholas Holyoak³ · Ranjodh B. Singh²

Received: 7 August 2019 / Accepted: 8 June 2020
© The Author(s) 2020

Abstract

A new method is proposed for generating projections for sparse populations by locality, age cohort and gender. An adaptation of the cohort replacement method, the approach uses a Tobit model with varying censoring limits to model population changes by cohort. As an applied example, projections are generated for 2016 Indigenous populations in communities in regional and remote Australia, and then compared to actual 2016 Census population counts. We argue the approach has considerable potential, with the model performing well in out-of-sample projection while offering projections at a much finer-grained level of disaggregation than currently available to planners and policy-makers.

Keywords Population modelling · Population projections · Cohort replacement · Regional and remote · Indigenous · Australia

JEL Classification C530 · J110

1 Introduction

Demographers and policy-makers are often faced with the challenge of projecting sparse populations in order to plan for the provision of infrastructure and basic services. This is especially the case in remote regions within a country. Without accurate spatial information regarding current and future population levels, it is almost

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10109-020-00329-z>) contains supplementary material, which is available to authorised users.

✉ Mark N. Harris
mark.harris@curtin.edu.au

¹ Bankwest Curtin Economics Centre, Curtin University, Perth, Australia

² School of Economics, Finance and Property, Curtin University, Perth, Australia

³ College of Science and Engineering, Flinders University, Adelaide, Australia

impossible to appropriately plan the provision of utilities and services such as water, housing, health and education. In this paper, we propose a new empirical approach to modelling sparse populations, illustrated through its application to projecting Indigenous populations of communities in regional and remote Australia.

By way of context, substantial disparities between Indigenous and non-Indigenous populations exist across virtually all indicators of social and economic well-being (Steering Committee for the Review of Government Service Provision 2016). The role of remoteness in narratives surrounding Indigenous disadvantage is both pivotal and controversial. Relative to other Australians, a far higher proportion of the Indigenous population—consisting of Aboriginal and Torres Strait Islander peoples—is found in remote areas of the continent. While large population centres offer greater access to services, infrastructure and labour market opportunity, for many Indigenous Australians those centres cannot offer the connections to homelands, culture and kinship networks intrinsic to their well-being.

Australia's current policy discourse includes reservations over the sustainability of remote Indigenous communities. Implicitly, and in some cases explicitly, it seems governments would prefer a rationalisation of these communities, with migration out of the smaller and more remote communities, and their eventual disappearance (see, for example, NT-Government 2009; Regional-Services-Reform-Unit 2016). Clearly the effectiveness of policy and funding allocations for remote communities relies upon a degree of consistency between assumed and actual population trends. That is, there is a paramount need for separate projections for the Indigenous population as a result of their vastly differing demography, service requirements and migration patterns relative to the mainstream population (Taylor et al. 2006; Wilson 2009; Raymer et al. 2017). Projections of the age structure of individual communities are important in determining the likely mix of services required, such as the number of school places and aged care facilities. As set out in the following section, such forecasting exercises are inherently difficult for sparse populations, with added complications in the context of Australia's Indigenous population. No projections currently exist at the community level to guide policy and planning.

2 Background

The Census of Population and Housing, conducted every 5 years by the Australian Bureau of Statistics (ABS), is the principal source of estimates of Australia's population and its demographic composition. It is also the main means by which the Indigenous population is counted, and since 1971, the intention has been for a full enumeration of that population (Wilson 2009). The 2011 Census recorded 548,000 people of Aboriginal or Torres Strait Islander descent, representing 2.5% of the total population. Table 1 shows how the Indigenous share of the population increases with remoteness. Only 34% of Indigenous Australians lived in major cities and 14% lived in areas classified as Very remote. In contrast, 71% of other Australians lived in major cities and less than 1% in areas classified as Very remote. So while Indigenous Australians made up around 2.5% of the overall population in 2011, they represented almost 41% of people living in Very remote Australia.

Table 1 Australian population estimates by Indigenous status and remoteness, 2011 census *Source:* ABS CData online table builder facility

Remoteness area (ARIA level)	Indigenous		Non-Indigenous		Total		Indigenous share
	People	%	People	%	People	%	
Major cities (1)	188,537	34.4	14,094,903	70.8	15,006,519	69.8	1.3
Inner regional (2)	121,301	22.1	3,695,423	18.6	3,998,424	18.6	3
Outer regional (3)	118,491	21.6	1,735,627	8.7	1,963,404	9.1	6
Remote (4)	39,751	7.2	234,833	1.2	300,107	1.4	13.2
Very remote (5)	77,493	14.1	98,564	0.5	190,266	0.9	40.7
Total	548,366	100	19,900,767	100	21,507,719	100	2.5

Columns sum to less than the total as the categories of migratory–offshore–shipping and no usual address have not been reported. Rows do not sum, as people for whom Aboriginal and Torres Strait Islander status was not stated have not been reported

One of the main approaches to population projection is the cohort component method (see Booth 2006 for a review). For given age cohorts in a base year, assumptions regarding deaths, births, immigration and emigration are applied to arrive at projections for that group at a future point in time. For a range of reasons, applying this method to project Indigenous populations in remote communities, which is done in this study, is fraught with additional complications likely to compound projection errors.

First, the accuracy of even the baseline data is questionable. There is evidence that earlier censuses undercounted young children and young to middle-aged adults, with inaccuracies more pronounced in remote areas (Taylor 1997). Despite considerable efforts on the part of the ABS, such problems persisted for more recent censuses (Wilson and Barnes 2007). In the 2011 Census, and again in 2016, there were twice as many people for whom Indigenous status was not stated as there were people who identified as being of Aboriginal or Torres Strait Islander descent, meaning error margins around those population estimates are very large.

The issue of identification creates a further source of population change within a cohort in addition to the three standard components of mortality, fertility and migration. This includes the question of how children from mixed families are likely to be identified, or identify themselves, in future censuses and the impact of policy changes on the propensity to identify (Taylor 1997; Wilson 2009; Biddle and Wilson 2013; Raymer et al. 2018). These challenges are specific to the enumeration of Indigenous Australians. On top of these, Taylor (2014) notes general problems associated with projecting populations for sparsely populated areas, including that they are more vulnerable to vagaries of exogenous impacts, such as weather events and policy changes; data collection is more resource-intensive; and proportional errors in projections tend to increase the smaller the population size of the units being analysed and if there is rapid change occurring in the period in which baseline data are compiled.

Further, most projection methodologies are based on the assumption of large population counts and cannot be applied to small populations. Taylor et al. (2006) suggest a regional population size of around 10,000 people is required to meaningfully apply age-conditional mortality analyses. Wilson (2011) suggests that exponential models have favourable properties over linear ones. However, if there are

zero counts in any component gender-by-age categories exponential models cannot be used, and inferences can be distorted by extreme values in terms of percentage change where counts are small.

An example of how these difficulties impact upon population forecasts for sparsely populated areas is provided by Taylor (2014)'s assessment of the accuracy of ABS projections for the Northern Territory from the 1970s through to 2012. Even at this territory level, Taylor (2014) found the mean absolute percentage errors in the ABS projections to be far higher than for Australia as a whole. Several relatively naive models, based on simple extrapolation of growth trends, outperformed the more sophisticated ABS cohort replacement model.

Most projections of the Indigenous population provide estimates at the national or state/territory level (see Wilson 2009). Taylor et al. (2006) nominate methodological developments in the treatment of small areas and subsequent small number analysis as a key imperative to improving demographic information for the Indigenous population in desert Australia. To the best of our knowledge, the smallest level at which projections of the Indigenous population have been provided in existing studies is from Biddle and Taylor (2009), covering 37 ABS-defined Indigenous Regions with as few as 2248 people, followed by projections from Taylor and Bell (2002) for the Cape York Peninsula with a baseline Indigenous population of 6500.

In contrast, the purpose of this research is to produce population projections for communities with as few as nine inhabitants and to apply this to remote Indigenous communities. This paper defines a modelling approach that has been developed specifically to deal with small population counts, and focuses on an intercensal (5-year) projection horizon. It thus aims to provide projections at a spatial level and time horizon crucial for policy formulation and planning, and where none currently exist. We note that although we are projecting Census estimates, this is of more policy relevance as it is exactly these numbers which form the basis for funding and other policy decisions.

3 Data

This study models intercensal changes based on the ABS-defined geography of Indigenous Locations:

Indigenous Locations (ILOCs) are aggregates of one or more Statistical Areas (Level 1). ILOCs generally represent small Aboriginal and Torres Strait Islander communities with a minimum population of 90 Aboriginal and Torres Strait Islander usual residents. An ILOC is an area designed to allow the production of census statistics relating to Aboriginal and Torres Strait Islander people with a high level of spatial accuracy while maintaining the confidentiality of individuals. For the 2011 Census, 1116 ILOCs have been defined to cover the whole of geographic Australia. There are non-spatial ILOCs for migratory–offshore–shipping and no usual address in each state and territory (ABS 2011).

The approach is to develop an empirical model of the 2011 Indigenous population counts for communities based on 2006 Census counts and selected characteristics

Census	Age					
j	1	2	...	15	16	17
2006	0–4 ↘	5–9 ↘	...	70–74 ↘	75+ ↘	
2011	0–4	5–9	...	70–74	75–79	80+

Fig. 1 Cohort structure of the census population data

of the communities. With the census undertaken every 5 years, and data available in 5-year age ranges, population changes for each community can be derived by age group. Census population estimates were downloaded from the ABS Table Builder online facility for all 1098 spatial ILOCs defined in 2011 and 838 spatial ILOCs defined for 2006. Data extracted for each ILOC include the population in 5-year age groups by sex and Indigenous status. An ABS-provided concordance was used to generate 2006 population estimates corresponding to the 2011 geography.

The resulting data include 2006 and 2011 population estimates for 1098 ILOCs that are geographically comparable between the 2 years. Census counts of the number of people identifying as Aboriginal, Torres Strait Islander or both Aboriginal and Torres Strait Islander were aggregated to a single Indigenous population estimate. For each ILOC in each census year, population data are available for 42 gender-by-age cohorts: 21 age groups (0–4, 5–9, ..., 95–99, 100+) each for males and females. For the older cohorts, a large proportion of these cells have either zero or very small population counts, and these were aggregated as shown in Fig. 1.

A key motivation was to generate projections for remote Indigenous communities. Given that population changes in the major cities and regional centres are likely to be driven by markedly different processes, only ILOCs in outer regional, Remote and Very remote Australia were included in the sample for estimation (corresponding to Accessibility/Remoteness Index of Australia [ARIA] levels of 3, 4 and 5). A small number of other ILOCs were excluded due to the fact that there were almost no Indigenous persons present at the time of the census. These exclusions were Lord Howe Island, Christmas Island, Cocos Islands and Apatula (Finke) Homelands and Walungurru Outstations (NT).

The final sample is based on data from 618 ILOCs, generating 19,776 observations for the regression analysis (32 observed cohort changes per ILOC). Also included in the model is the natural logarithm of the total population of the ILOC (including non-Indigenous people) to capture differences by community size; dummy variables for cohort age, gender, ARIA and state/territory; the gender- and age-specific 5-year survival rate for each cohort; and interaction terms between age group and ILOC size, and between age group and ARIA.

It is worth mentioning that census data used in this analysis have been checked for quality. Two measures of data quality in a census are response rates and undercounting. For the 2016 census, the response rate was 95.1% and the net undercount was 1%. This is a good result when compared to other countries such as Canada, UK and New Zealand. With regard to Indigenous populations, *The Census is the only comprehensive source of small area data on the Aboriginal and Torres Strait Islander population* (ABS 2018).

4 Method

Following the cohort component approach, note that the number of Indigenous persons aged 5–9 in a given ILOC in 2011 will be the number who were aged 0–4 in 2006 minus deaths, net migration (includes internal migration) and net changes associated with identification of Indigenous status.¹ The data set-up is demonstrated in Fig. 1, where the arrows trace the cohort movement through time. Using j to index the age group categories, it can be seen that there are 16 such cohort progressions or flows in which the population (P) in 2011 (t) can be related to the population in the earlier age category in 2006 ($t - 1$):

$$P_{j-1,t-1} \rightarrow P_{jt}. \quad (1)$$

With population data for males and females, this gives 32 flows observed for each ILOC. This can be treated as a multi-level regression framework in which there are 32 observations for each ILOC. Following the convention in the econometric multi-level modelling literature (Rabe-Hesketh and Skrondal 2008, p. 65), the entity or cluster is denoted by subscript i and the occasions providing repeat observations for that cluster by subscript j . Hence, in this case communities are denoted by subscript i ($i = 1-618$) and the 2011 age groups by subscript j ($j = 2-17$). For convenience, the gender distinction is ignored for the purposes of setting out the model. From Eq. 1, a modelling framework can be developed which incorporates clustering at the ILOC level,

$$P_{ijt} = f(P_{ij-1,t-1}). \quad (2)$$

At this stage, the cohort model does not take into account any changes in population due to births, deaths, migration and identification issues of Indigenous individuals. However, clearly all of these components will affect observed population levels. In the empirical approach described more fully below, we have an explicit fertility model to account for the former, and we include death rates in the model to address the second issue. However, it is not clear how to address the final two points in the empirical strategy, and therefore, we make the assumption that these rates are relatively constant over time, such that when we model *changes* (see below) in population levels, such systematic biases in the population level counts are effectively “differenced out”.

Here, we suggest that the functional relationship between P_{ijt} and $P_{ij-1,t-1}$ can be estimated taking into account both observable and unobservable ILOC characteristics and effects. A range of options are available for modelling changes in cohort populations between 2006 and 2011. Exploration of these indicated that it is preferable to model changes in linear terms (or levels) rather than in growth terms. Models based on proportionate change are highly sensitive to extreme values that mostly arise where cell counts are small. For example, percentage changes in some age categories in smaller communities were in the thousands and are likely to be affected

¹ It is assumed that the net migration rates are constant over the short to medium term.

by concordance and enumeration issues. Modelling changes in growth rates also means omitting observations for which the base number is zero, even though that is a legitimate and common value.

However, modelling current population levels as a function of past levels, especially over a short number of time observations, raises several econometric concerns, such as endogeneity, non-stationarity and the spurious regression problem. Indeed, in a simple regression, the lagged dependent variable clearly exhibited signs of these issues, with the estimated coefficient being very close to unity in value and with an extremely high t -statistic. Thus, the preferred strategy is to model the change in population for each cohort. That is, the dependent variable becomes:

$$c_{ijt} = P_{ijt} - P_{i,j-1,t-1}. \quad (3)$$

As such there are 32 observations of c_{ijt} for each ILOC ($i = 2$ to 17 for males and females). Referring back to Fig. 1, the model estimates changes in the population of people aged 5–9 in 2011 from the number aged 0–4 in 2006; of people aged 10–14 in 2011 from the number aged 5–9 in 2006; and so on. However, it cannot provide estimates for the population aged 0–4 in 2011, as there is no younger cohort in 2006 to use as the baseline. To enable projections for the total populations by ILOC, a separate fertility model is developed to generate estimates of the number of males and females aged 0–4 in 2016, which we detail below.

4.1 Apparent fertility rates

Regarding fertility trends for Indigenous Australians, these have been decreasing overall for both remote and non-remote regions. Additionally, the biggest drop has occurred in the 15–19 age group. Researchers have attributed this change to increased educational attainment (Venn and Dinku 2019). The evidence also seems to suggest that the Indigenous fertility rates are converging to non-Indigenous rates over time. Based on this, we attempt to model the fertility component of the proposed population model.

To develop a model to predict the Indigenous population aged 0–4 in each ILOC, a linear regression model was estimated across ILOCs with the 2011 Indigenous population aged 0–4 as the dependent variable. Models were tested with a variety of specifications that included summations of the male and female Indigenous populations in the ILOC in 2006, initially focusing on what were considered to be adults of childbearing age. However, experimenting with which age groups to include and allowing differential effects by age and by region to maximise the model fit returned the relatively simple model set out in Eq. (4). A model that included the number of Indigenous children aged 0–14 in the ILOC in 2006 interacted with ARIA dummies was found to have the best predictive capability. For the full sample of ILOCs across all of Australia, the model returned an adjusted R-squared of 0.9599.

The number of Indigenous males aged 15–34 marginally added to the predictive power, but this component of the model applied only to projections for ILOCs in the major cities (i.e. ARIA level 1). For ILOCs in outer regional, Remote and

Very remote Australia, the predicted number of Indigenous people aged 0–4 years is given by:

$$P_{1,j,t} = -2.37 + 0.42 * OREG_j \sum_{i=1}^3 P_{i,j,t-1} + 0.35 * REM_j \sum_{i=1}^3 P_{i,j,t-1} + 0.35 * VREM_j \sum_{i=1}^3 P_{i,j,t-1}, \quad (4)$$

where P_1, P_2 and P_3 denote the Indigenous population aged 0–4, 5–9 and 10–14, respectively. OREG, REM and VREM are dummy variables indicating that the ILOC is in outer regional, Remote and Very remote Australia. In the estimated model of the 2011 population on 2006 values, each of the estimated coefficients (including the intercept term) was significant at the 1% level. The coefficients were applied to the 2011 data to obtain projections for the population aged 0–4 in 2016 for each ILOC in our analysis sample. The projected population was allocated as 50% boys and 50% girls. The cases where the predicted population was negative, the projection was set to zero.

4.2 Statistical model

Stochastically then, population changes can be modelled as:

$$c_{ijt} = \mathbf{x}'_{i,j,t-1} \boldsymbol{\beta} + u_i + \lambda_j + \varepsilon_{ijt} \quad \text{where} \quad \varepsilon_{ijt} \sim N(0, \sigma^2) \quad (5)$$

and $\mathbf{x}_{i,j,t-1}$ represents a vector containing the independent variables (all defined in the base year 2006); $\boldsymbol{\beta}$ is a vector of coefficients to be estimated; u_i and λ_j denote the unobservable community-level (ILOC) and cohort effects. Lastly, ε_{ijt} denotes the usual errors in the model.

The above model (Eq. 5) can be estimated by simple linear regression. However, this is not appropriate for the “small numbers” model we are dealing with here, as the data will be necessarily truncated in many instances as the dependent variable has an effective lower bound. That is, any population cannot decrease by more than the starting value. For example, consider an ILOC with five Indigenous males aged 20–24 in 2006. The change in the population from 2006 to 2011 can only range from –5 upwards, given that the population of males aged 25–29 in 2011 cannot be less than zero.

Note also that if the population of males aged 25–29 in 2011 is also five, then there has been no change and the dependent variable c_{ijt} equals zero. This zero is a legitimate value indicating no change in the population, and indeed in any estimation, the expected value of c_{ijt} would (should) accordingly be close to zero. However, consider the case in which there are no individuals in a particular age-by-gender category in 2006, as is common for older age cohorts. In this case, the population can increase ($c_{ijt} > 0$), but has a lower bound of zero, and the probability of observing zero change is much higher than for observations with a positive initial population. In such situations where there is a ‘latent’ potential change in the population that cannot be observed because of the effective lower bound, linear regression models will produce biased and inconsistent results, which will

worsen with the extent of such censoring/boundary observations (Amemiya 1984; Greene 2003).

A preferable approach is to implement a Tobit model with varying censoring limits. As noted above, in the current context the limit is equal to the 2006 population for a specific gender-by-age group at a given ILOC. This differs from the usual Tobit model set-up, where lower (and/or upper) limits are usually fixed (commonly at zero) and the same for all observations in the sample.

To be clear, a standard Tobit model typically has a fixed lower (upper) limit, which is often a lower bound at zero. However, in the suggested approach, as alluded to above, the situation here is subtly, but importantly, different. With a starting population of X (with $X \geq 0$), then (the negative of) this provides an effective lower bound for the change in the next period: by definition, the change cannot be $-Y$ where $|Y| > X$ as it is not possible to have negative actual population levels. That is, while population *changes* can lie anywhere on the real number line, population *levels* must be strictly ≥ 0 . However, this latter condition has direct implications for the allowable range of the former. Unfortunately, the situation is somewhat complicated here as in the cohort component model employed—as defined by Eqs. (1) and (2)—the cohort necessarily ages from $j - 1$ to j , as in equation (3). In this way, each c_{ijt} observation faces an effective, and binding, lower bound of not $-P_{i,j,t-1}$, but $-P_{i,j-1,t-1}$, due to the necessary ageing of the cohort. Thus, the usual fixed lower limit Tobit model has to be amended to an observation-varying one defined by the variable $-P_{i,j-1,t-1}$.

The previous model (Eq. 5) can be updated to reflect this possible censoring as

$$c_{ijt}^* = \alpha_i + \mathbf{x}'_{i,j,t-1} \beta + \varepsilon_{ijt} \quad \text{where} \quad \varepsilon_{ijt} \sim N(0, \sigma^2) \tag{6}$$

and c_{ijt}^* now denotes the *latent* underlying change in ILOC i for age group j at time t (2011). However, this cannot be fully observed due to the fact that the change cannot be less than the current population. In other words, there is lower tail censoring such that only c_{ijt} is observed. Hence,

$$\text{if } c_{ijt}^* < -P_{i,j-1,t-1} \quad \text{then} \quad c_{ijt} = -P_{i,j-1,t-1}. \tag{7}$$

In contrast to the standard Tobit model in which the lower limit is assumed to be a fixed value for all i and j , the proposed framework contains the 2006 population as a varying lower limit for each observation. Like the linear model, the Tobit model can be estimated assuming either random or fixed effects, although the latter will suffer from the well-known incidental parameters problem, if the dimension over which these are constant is 'small' (Greene 2012). Additionally, estimating fixed effects for a large number of ILOCs (over 600) is problematic. With regard to modelling cohort effects, these were incorporated into the explanatory variables. Hence, after extensive modelling, it was established that the Tobit model with random effects provided a better fit.

The data used for developing the model contains only one observation on each c_{ijt} , i.e. the change from 2006 ($t - 1$) to 2011 (t). In this sense, the model is cross-sectional rather than longitudinal, and as such, the time subscript can be omitted. With 2016

Census data now available, further work is planned to move to a true multi-level panel structure that should provide more rigorous estimation of community-level unobserved effects. For now, the random-effects Tobit model can be expressed as

$$c_{ij}^* = \mathbf{x}_{ij}' \boldsymbol{\beta} + \varepsilon_{ij} + u_i. \tag{8}$$

Following usual practice, the identifying assumption is that ε_{ij} and u_i are both normally distributed, with zero means and variances of σ^2 and ω^2 , respectively. The data are observed as $c_{ij} = \max(L_{ij}, c_{ij}^*)$ where $L_{ij} = -P_{i,j-1}$. In this context, this is an example of lower tail censoring: the change in population in the next time period cannot be less than the number of people (in an age group) currently residing in the ILOC. As per usual, the ε_{ij} is assumed to be uncorrelated across ILOCs. To derive the log likelihood function, the focus here is on the conditional distribution of $f(c_{ij}|u_i)$. Let the dummy variable, $d_{ij} = 1$ indicate that $c_{ij} > L_{ij}$. This is the uncensored case and $d_{ij} = 0$ for censored cases. Based on the above identifying assumptions, the conditional density of c_{ij} can then be expressed as

$$f(c_{ij}|u_i, d_{ij} = 0) = P(c_{ij}^* \leq L_{ij}|u_i) = \Phi\left(\frac{L_{ij} - \mathbf{x}_{ij}' \boldsymbol{\beta} - u_i}{\sigma}\right)$$

for censored cases and

$$f(c_{ij}|u_i, d_{ij} = 1) = \frac{1}{\sigma} \phi\left(\frac{c_{ij} - \mathbf{x}_{ij}' \boldsymbol{\beta} - u_i}{\sigma}\right)$$

for uncensored cases (Greene 2012), where Φ and ϕ , respectively, denote the *c.d.f* and *p.d.f* of the standardised normal distribution. Combining the above two cases,

$$f(c_{ij}|u_i) = [f(c_{ij}|u_i, d_{ij} = 0)]^{1-d_{ij}} \times [f(c_{ij}|u_i, d_{ij} = 1)]^{d_{ij}}.$$

Assuming independence, the joint density of all observations in a group can be expressed as

$$f(c_{i1}, c_{i2}, \dots, c_{iT_j}|u_i) = \prod_{j=1}^{T_j} f(c_{ij}|u_i).$$

Based on the above results, the log likelihood function of this model can be written as

$$\log L = \sum_{i=1}^n \log \left\{ \int_{-\infty}^{\infty} \frac{1}{\omega \sqrt{2\pi}} \exp\left(-\frac{u_i}{2\omega^2}\right) \prod_{j=1}^{T_j} \left[\Phi\left(\frac{L_{ij} - \mathbf{x}_{ij}' \boldsymbol{\beta} - u_i}{\sigma}\right) \right]^{1-d_{ij}} \left[\frac{1}{\sigma} \phi\left(\frac{c_{ij} - \mathbf{x}_{ij}' \boldsymbol{\beta} - u_i}{\sigma}\right) \right]^{d_{ij}} du_i \right\}.$$

Lastly, find values of β , σ and ω such that this function is maximised. The integrals can be computed using Gauss–Hermite quadrature (or by simulation) and the function can be maximised using standard nonlinear optimisation methods.²

We note that although the above statistical models are rather complex, they can be estimated routinely in standard commercial software. For example, here we used the *Limdep/Nlogit version 6* package, although *Stata version 16* could also be used. For these reasons, the suggested approach is easy to use and implement, and could therefore be widely applicable to any other research areas characterised by sparse populations.

In order to model change, several variables were considered. For example, a priori one can expect the initial population size of the ILOC to be a factor that affects change. In addition to this, age group, remoteness level, state and gender are also factors. Furthermore, the interaction terms are included to allow for possible differential effects of age by ILOC size and remoteness. Previous studies have identified trends in which younger Indigenous people tend to move away from smaller, more remote communities into larger regional centres, while older people tend to move back to country (Biddle 2009). As such, we tested all available variables and various interaction terms to allow for a flexible specification. The results indicated that interaction terms were clearly preferred since majority of them were significant. For example, in the case of ILOC size and age group, 11 of interaction terms are significant out of 15 possible terms.

A dummy variable was also included to indicate whether the community was nominated as a Territory Growth Town under the Northern Territory Governments' 2009 Working Future policy (Sanders 2010).

Also included in the model as a covariate are survival rates for each age group (taken from separately available ABS projections of the total Indigenous population by age). As expected, these rates are an important factor for modelling population change. The survival rates are calculated for each gender and age group. It is the ratio of the number of individuals in age group i at time t to the number of individuals in age group $i - 1$ at time $t - 1$. The survival rates are close to unity for younger cohorts and decline to under 0.7 for cohorts beyond the age of 70 years. Not surprisingly, this variable is highly significant in the final model. Note that we assume the survival rates for each gender and age group are constant over the short to medium term. For further details on all the variables included in the model, please refer to Table 2. Descriptive statistics for all variables are provided in "Appendix 1".

5 Results

The model coefficients and standard errors of the Tobit model with varying lower limits are provided in "Appendix 2". Note that these results pertain to the pooled version of the model. Having conditioned on all the observed heterogeneity, there seemed to be very little remaining unobserved heterogeneity

² The results for these are available upon request.

Table 2 Definitions of variables in the data set

Variable(s)	Description
ILOC size	The natural logarithm of the resident total population of the ILOC in 2006 (including Indigenous, other Australians and those for whom Indigenous status is not stated)
Remote, Very remote	Two mutually exclusive dummy variables indicating whether the ILOC is in ARIA level 4 (Remote) or ARIA level 5 (Very remote). The omitted or 'reference' category is outer regional (ARIA level 3)
Victoria, Queensland, South Australia, Western Australia, Tasmania, Northern Territory	Six mutually exclusive dummy variables indicating the state or territory of the ILOC. New South Wales is the omitted category. There are no outer regional, Remote or Very remote ILOCs in the Australian Capital Territory
Female	Dummy variable is equal to 1 if the observation is for a female cohort and is equal to 0 if it relates to a male cohort
Growth town	Dummy variable taking on a value of 1 if the ILOC contains or corresponds to one of the communities nominated as growth towns under the Northern Territory Government's <i>Working Future</i> policy announced in 2009. While the policy named 20 towns, one of these (Daguragu–Kalkarindji) falls across two ILOCs (Daguragu and Kalkarindji), meaning there are 21 ILOCs coded with a value of 1
Age 10–14; Age 15–19, ... Age 75–79, Age 80+	Sixteen mutually exclusive dummy variables indicating the age of the cohort in 2011. The omitted category is age 5–9
Survival rate	Based on ABS Catalogue 3238.0— <i>Estimated and projected Aboriginal and Torres Strait Islander population Series B for Australia</i> (ABS 2014). The ratio of the estimated population in each age cohort i in 2011 to the estimated population of age cohort $i - 1$ in 2006. This gives an age-specific <i>apparent survival rate</i> and is calculated separately for males and females
ILOC size*age interaction terms	Fifteen separate variables are generated by interacting the continuous ILOC size variable with the age group dummies. The omitted age category is age 5–9. The coefficients on these variables indicate whether, within each specific age group, there is any further effect of community size in addition to the average effect of community size observed across all age cohorts

Table 2 (continued)

Variable(s)	Description
Outer regional \times age and Remote \times age interaction terms	Thirty separate dummy variables generated by interacting the outer regional dummy variable (ARIA level 3) with age cohort and the Remote dummy variable (ARIA level 4) with age cohort. The omitted categories are Very remote (ARIA level 5) and Age 5–9. The coefficients on these variables indicate whether, within each specific age group, there is any further effect of remoteness in addition to the average effects observed across all age cohorts

(at the ILOC level) such that statistically the pooled version of the Tobit model was preferred. The model specification consisted of ILOC size, survival rate for each age cohort, dummy variables for gender, level of remoteness, state and age cohort (to capture heterogeneity across age groups). Also included in the model were interaction terms between age cohorts, ILOC size and remoteness levels. In addition to this, a policy variable (growth town) was also included in the model. Given these variables, the model's primary purpose was to generate population projections. Based on the results presented, the significant drivers of the model include ILOC population size, survival rate, some of the age cohort dummies and most of the interaction terms.

The proposed Tobit model does perform better than a simpler specification such as a linear regression model. The added feature of censoring through varying lower limits enables the proposed model to always produce meaningful predictions. A comparison between the proposed model and the simple model was made. This was done by comparing the predictions resulting from both models to actual values. The correlation coefficient for actual versus fitted values from the simple model was 23% versus 33% for the proposed model. Additionally, the simple model predicts nonsensical results for 15% of the observations. This implies that the change in population numbers for a given age group is larger than the censored value. As such, the simple model predicts a change that is not possible. The proposed model accommodates the censored value and hence produces a meaningful prediction. Indeed, for observations where the change in population is equal to their lower limit, i.e. censored cases (30% of all cases), the correlation between actual versus fitted is 61% for the proposed model versus 3% of the simple model. Thus, the proposed model clearly provides a significantly better specification.

6 Out-of-sample Projections

The development sample for the proposed model was 2006 to 2011. The resulting model was used to conduct in-sample testing as well as out-of-sample testing. Both in-sample and out-of-sample testing shows that the proposed model can capture trends across one and two census horizons. In-sample testing consists of simply

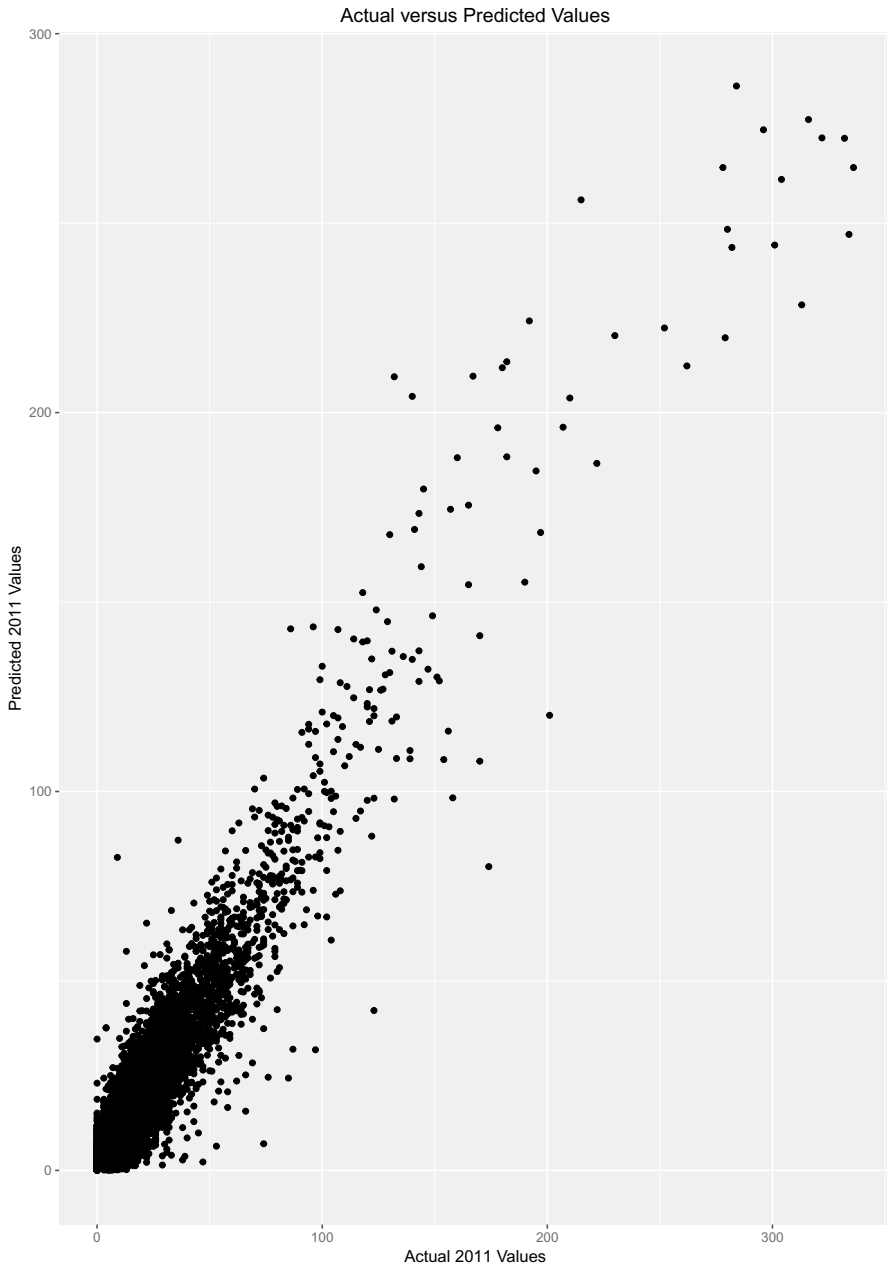


Fig. 2 Actual 2011 values versus predicted 2011 values

adding the predicted change to the actual 2006 values to obtain the predicted population numbers for 2011 and then comparing these values to actual population numbers for 2011 for all age groups across all ILOCS. Figure 2 shows the actual 2011

Table 3 Indigenous populations: 2006 and 2011 census estimates and 2016 projections, by remoteness

ARIA	No. of ILOCs	Population		Pop. growth (%)		
		Actual	Predicted	2006–2011 Actual	2011–2016 predicted	2011–2016 actual
3-Outer regional	245	126,142	139,953	19.4	19.7	7.9
4-Remote	95	38,412	43,073	9.5	10.6	– 1.3
5-Very remote	278	75,947	84,384	7.1	9.3	– 1.6
Total	618	240,501	267,409	13.4	14.8	3.2

values versus predicted 2011 values produced by the Tobit model for all 19,776 ILOC-by-gender-by-age cases.

The critical test of the methodology is how well it performs in out-of-sample prediction of the Indigenous populations. Recall the model is developed using only data from the 2006 and 2011 Censuses. Applying the estimated coefficients from the cohort and fertility models to the 2011 data, including estimated ILOC-specific effects, generates projections of the 2016 populations by gender and 5-year age group for every ILOC. These can simply be summed to produce predictions of the total 2016 population for each ILOC, or to other aggregated levels. The release of the 2016 Census by Indigenous geography subsequent to the model development provides actual population outcomes against which to compare the model predictions.³

As set out above, the model has been developed to enable use of panel techniques. In its current application, however, we have only one observation on population changes (the c_{ijt} s). So while we exploit the availability of data in two time periods to calculate population changes between 2006 and 2011, the estimation is a cross-sectional one. In our data, the aggregate Indigenous population across the regional and remote ILOCs increased by 13.4 per cent from 2006 to 2011. By construction, the model projects forward broadly similar growth from 2011 to 2016, at 14.8 per cent. The actual change in the Census count of the Indigenous populations from 2011 to 2016 turned out to be an increase of just 3.2 per cent. Hence, at the aggregate level, we over-predict the population. However, comparison of actual versus predicted outcomes shows that the model performs extremely well in picking up key trends. Table 3 shows that this over-prediction at the aggregate level is evident at each ARIA level, but the general pattern is replicated with population growth predicted to be strongest in outer regional Australia and lower in Remote and Very Remote Australia, with the latter the lowest by a small margin.

³ There were a number of changes to the ILOC boundaries between the 2011 and 2016 Censuses. In two cases, two 2011 ILOCs were amalgamated to a single ILOC in the 2016 data, while in two other cases the reverse occurred—a single 2011 ILOC was split into two 2016 ILOCs. Either the baseline data or projections were aggregated to ensure comparability of actual and predicted populations. However, adjustments have not been made for numerous fractional concordance changes for ILOCs existing in both the 2011 and 2016 data.

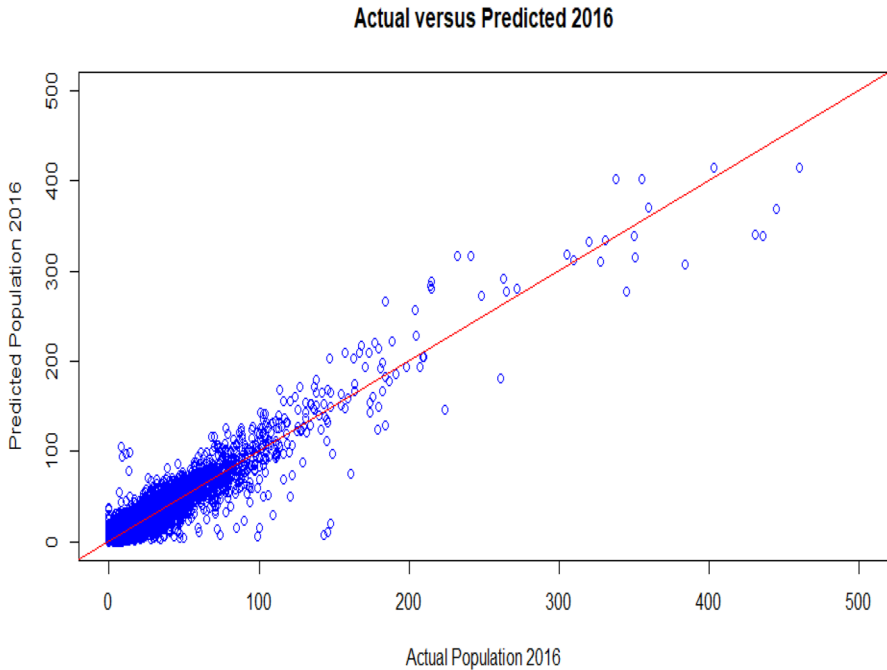


Fig. 3 Actual 2016 values versus predicted 2016 values

For the individual ILOC-by-gender-by-age cells, the average difference between the actual and predicted population values is -1.37 persons, reflecting the noted over-prediction of populations in aggregate. The correlation between actual and predicted values is approximately 0.95. As can be seen in the actual versus predicted plot (Fig. 3), the model predictions are around the 45-degree line (shown in red), suggesting minimal systematic bias.

More importantly, the model does a remarkable job in accurately predicting small populations. In fact, there are seven cases where the actual outcome is equal to zero and the model predicts exactly that. Furthermore, if the least accurate 1% of the predictions are removed, the maximum and minimum differences between actual versus predicted outcomes improve significantly: the minimum error increases from -96 to -25 and the maximum error drops from 134 to 25.

When the projections are aggregated to the ILOC level (Fig. 4), the correlation between the actual and projected populations is 0.983, indicating the model is extremely robust in forecasting community population levels over a 5-year horizon.

In terms of the changes in ILOC populations between 2011 and 2016, the correlation between the actual and predicted changes is 0.49 and highly significant ($p < 0.0001$). Closer inspection reveals encouraging forecasting successes given the random nature of population changes in remote communities. Out of the 616 ILOCs for which actual and predicted comparisons can be made, the largest actual increase occurred in Thuringowa, a suburban area of Townsville in regional northern Queensland, with growth in the Indigenous population of 1605 persons. Remarkably, this

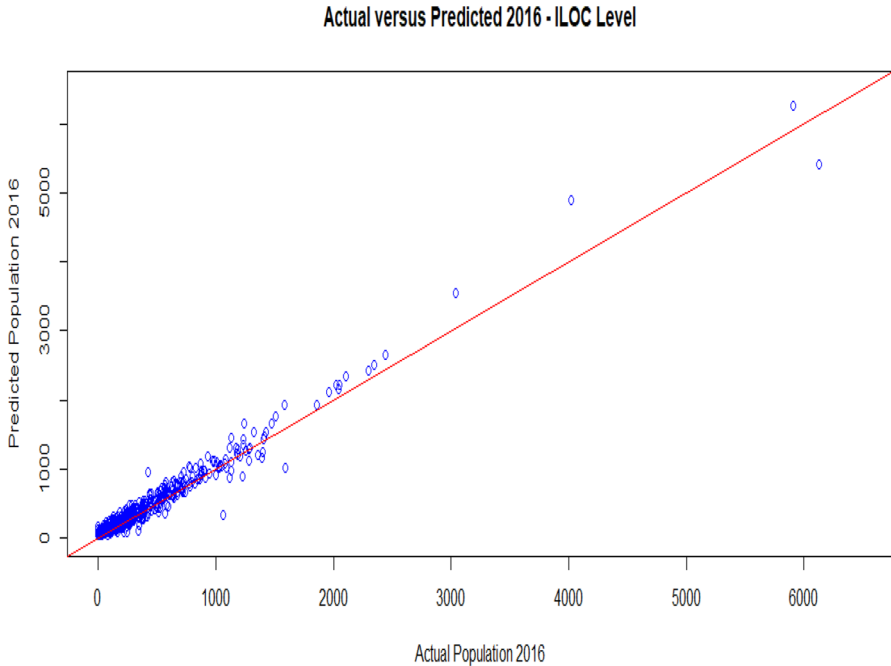


Fig. 4 Actual 2016 values versus predicted 2016 values at ILOC level

is also the ILOC projected by the model to have the largest increase, although it substantially underestimates the increase at 888 persons. The ILOC projected to have the largest decrease in population was remote Miali Brumby–Warlpiri on the outskirts of Katherine in the Northern Territory. The Indigenous population was projected to decline by 38 persons, compared to an actual decline of 44 persons as recorded in the census. However, it must be acknowledged that a number of ILOCs experienced much larger declines that were not predicted by the model.

Finally, Fig. 5 shows the actual and projected Indigenous population for regional and remote Australia when the ILOC estimates are aggregated to 5-year age groups by gender. The structure of the age pyramid is well captured by the projections, with the exception of the discord between the 2016 actual and projected number of males in the oldest category. Indeed, in this case the actual Census figures seem somewhat anomalous to the baseline data, with the number of males aged 75 and over in 2016 being around two-thirds the number in the 70–74-year age group. In previous years, the count of males aged 75+ was in fact larger than the number aged 70–74 years.

The primary policy variable included in the analysis was the indicator of whether the ILOC included one of 20 towns nominated as a Northern Territory Growth Town. As reported in “Appendix 2”, the coefficient on this variable was negative and insignificant, meaning that these towns were associated with population changes between 2006 and 2011 no different to what would have been expected given their characteristics. Accordingly, the 2016 census indicates that the growth in the population of these ILOCs, at 2.7 per cent, grew marginally slower than the overall

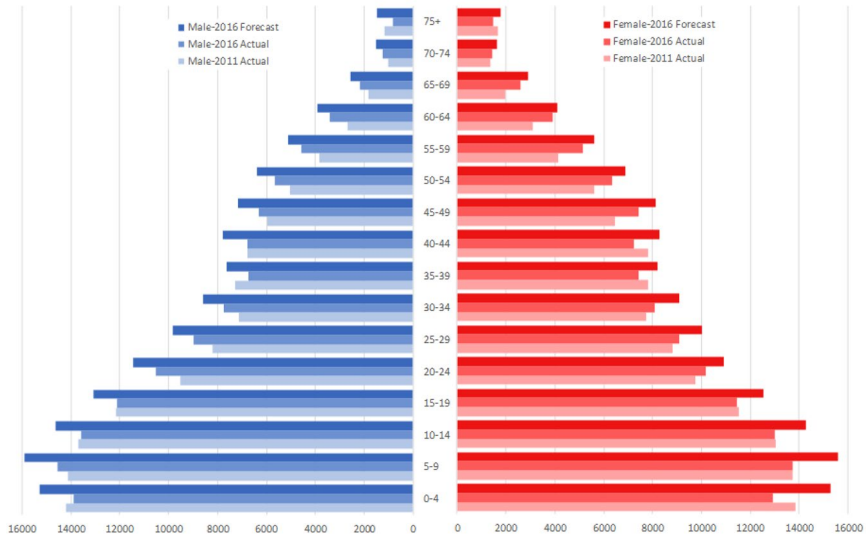


Fig. 5 Age structure of the Indigenous population by gender: actual 2011 and 2016 and forecast 2016

average. While we do not intend to put great store in this result, it demonstrates how variables can be readily incorporated into the framework to assess the impact of policy measures.

With regard to longer-term projections, based on our experience, governments at all levels (local, state and federal) as well as other policy-makers are all focussed on short to midterm population projections. This information is used to assess the feasibility of proposed regional developmental programs. Long-term population projections appear to be rarely useful in this context. Furthermore, almost all models will produce inaccurate long-term forecasts. As such, we produce projections for one and two census horizons.

7 Discussion

This paper proposes an approach for generating population projections at a level of spatial and demographic detail that would typically be precluded using existing methods because of the small numbers involved. The approach is a variant of the cohort replacement model, but focuses on changes in population levels within a regression framework. The focus on changes in population levels negates difficulties associated with models based on growth rates when handling small and zero population counts. The regression model specification—a Tobit model with varying censoring limits—ensures the regression estimates are not biased due to ignoring the censored observations, while at the same time eliminating the possibility of the model projecting negative populations. The model can be used to generate detailed projections and provide additional information of relevance to policy-makers from

the effects implied by regression coefficients, such as characteristics associated with fast-growing or declining areas.

The approach was developed and used to generate projections of Indigenous populations in remote communities in Australia. We use that work as an example of the application of the methodology and to test its performance although the approach may be applied to other such characterised population profiles, depending on the availability of necessary data and scope of application. The techniques, although quite complex in nature, can be applied using standard statistical/econometric packages, making them easy to use and therefore widely applicable across a broad range of end-users. Projections of community populations for 2016 by age and gender are generated from a model developed with 2006 and 2011 Census data, and the robustness of those projections tested through comparison with subsequently released 2016 Census data. Given the high degree of randomness in population changes in such communities and extensive concordance adjustments required to align the 2006 and 2011 data, we believe the model is shown to perform remarkably well in out-of-sample projection, with some scope for further enhancement of the model.

The proposed model was compared to a simple model and it was shown that the proposed model clearly outperforms the latter one. Next, the predictions from the proposed model were subject to in-sample as well as out-of-sample testing. In both cases, the proposed model performed remarkably well as shown in Figs. 2 and 3. There are, however, certain factors pertaining to both the data and model which could impact the projections. These include issues regarding data accuracy (undercounting Indigenous populations in the census and identification of Indigenous individuals) as well as model assumptions on fertility, net migration and death/survival. We have attempted to address each of these in most pragmatic way possible.

As noted by Wilson (2011): *Probably the logical starting point in the design of any projection system is to consider its purpose ...what question or problems does it need to solve? What projections are required to solve these problems? And of the required outputs, what is the most important and should be prioritised given the resources available?* Our application responds to a pressing need for sparsely populated area estimates for practical purposes of policy and planning for remote Aboriginal communities and for methodological advances in small number analysis more generally to meet the demographic informational needs of Indigenous Australians. Existing projections of Australia's Indigenous population disaggregated by detailed age group are available on a regular basis only at the level of the 8 states and territories, a far cry from the community level projections possible with this approach.

While demographic projections are often prepared 20, 30 or even 50 years into the future, this paper has focused on generating projections 5 years ahead of a baseline census, namely projecting 2016 populations based on 2011 data. It is quite straightforward to use the method to project to 2021 and beyond, by taking those 2016 projections as the baseline and reapplying the model coefficients. Of course, projection errors will increase with the forecasting horizon and we would have no way yet of testing the accuracy of those projections. In any case, a projection horizon of around 5 years seems more in line with policy and planning cycles impacting upon the Indigenous communities in regional and remote Australia, and we suspect that would be true of many other potential applications for

this methodology. The approach is also very economical in its resource requirements, requiring essentially only past population data by region and age. In our example, almost all the input data came from the ABS Census and are publicly accessible from the ABS website.

Most importantly, while we have estimated a cross-sectional model, the framework was conceived to be longitudinal in nature. We anticipate that for most potential applications analysts would have more extensive time series data available and could therefore model additional dimensions of unobserved heterogeneity.

Finally, as is always the case with such projections, we caution that estimates for small areas should be interpreted with appropriate recognition of their limitations. Projections based on modelling results for local communities should only be used in decision-making following corroboration through alternative data sources, local knowledge and local consultation.

Acknowledgements The work reported in this publication was supported by funding from the Australian Government Cooperative Research Centres Program (CRC) through the CRC for Remote Economic Participation (REP). The views expressed herein do not necessarily represent the views of the CRC-REP or Ninti One Limited or its participants. Errors or omissions remain with the authors.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix 1: Descriptive statistics for all variables included in the regression model ($n = 19,776$)

Variable	Mean	SD	Min.	Max.
Population change (dep. var.)	- 0.0295	6.0961	- 79.71	100.78
ILOC size (log)	6.9704	1.6393	2.17	11.36
Remote	0.1537	0.3607	0	1
Very remote	0.4498	0.4975	0	1
Victoria	0.0162	0.1262	0	1
Queensland	0.2104	0.4076	0	1
South Australia	0.0744	0.2625	0	1
Western Australia	0.2120	0.4087	0	1

Variable	Mean	SD	Min.	Max.
Tasmania	0.0356	0.1853	0	1
Northern Territory	0.2961	0.4566	0	1
Female	0.5000	0.5000	0	1
Growth town	0.0340	0.1812	0	1
Age 10–14	0.0625	0.2421	0	1
Age 15–19	0.0625	0.2421	0	1
Age 19–24	0.0625	0.2421	0	1
Age 25–29	0.0625	0.2421	0	1
Age 30–34	0.0625	0.2421	0	1
Age 35–39	0.0625	0.2421	0	1
Age 40–44	0.0625	0.2421	0	1
Age 45–49	0.0625	0.2421	0	1
Age 50–54	0.0625	0.2421	0	1
Age 55–59	0.0625	0.2421	0	1
Age 60–64	0.0625	0.2421	0	1
Age 65–69	0.0625	0.2421	0	1
Age 70–74	0.0625	0.2421	0	1
Age 75–79	0.0625	0.2421	0	1
Age 80+	0.0625	0.2421	0	1
Survival rate	0.9012	0.1252	0.60	1.00
ILOC size × Age 10–14	0.4356	1.7364	0	11.36
ILOC size × Age 15–19	0.4356	1.7364	0	11.36
ILOC size × Age 19–24	0.4356	1.7364	0	11.36
ILOC size × Age 25–29	0.4356	1.7364	0	11.36
ILOC size × Age 30–34	0.4356	1.7364	0	11.36
ILOC size × Age 35–39	0.4356	1.7364	0	11.36
ILOC size × Age 40–44	0.4356	1.7364	0	11.36
ILOC size × Age 45–49	0.4356	1.7364	0	11.36
ILOC size × Age 50–54	0.4356	1.7364	0	11.36
ILOC size × Age 55–59	0.4356	1.7364	0	11.36
ILOC size × Age 60–64	0.4356	1.7364	0	11.36
ILOC size × Age 65–69	0.4356	1.7364	0	11.36
ILOC size × Age 70–74	0.4356	1.7364	0	11.36
ILOC size × Age 75–79	0.4356	1.7364	0	11.36
ILOC size × Age 80+	0.4356	1.7364	0	11.36
Outer regional × Age 10–14	0.0248	0.1555	0	1
Outer regional × Age 15–19	0.0248	0.1555	0	1
Outer regional × Age 19–24	0.0248	0.1555	0	1
Outer regional × Age 25–29	0.0248	0.1555	0	1
Outer regional × Age 30–34	0.0248	0.1555	0	1
Outer regional × Age 35–39	0.0248	0.1555	0	1

Outer regional × Age 40–44	0.0248	0.1555	0	1
Outer regional × Age 45–49	0.0248	0.1555	0	1
Outer regional × Age 50–54	0.0248	0.1555	0	1
Outer regional × Age 55–59	0.0248	0.1555	0	1
Outer regional × Age 60–64	0.0248	0.1555	0	1
Outer regional × Age 65–69	0.0248	0.1555	0	1
Outer regional × Age 70–74	0.0248	0.1555	0	1
Outer regional × Age 75–79	0.0248	0.1555	0	1
Outer regional × Age 80+	0.0248	0.1555	0	1
Remote × Age 10–14	0.0096	0.0975	0	1
Remote × Age 15–19	0.0096	0.0975	0	1
Remote × Age 19–24	0.0096	0.0975	0	1
Remote × Age 25–29	0.0096	0.0975	0	1
Remote × Age 30–34	0.0096	0.0975	0	1
Remote × Age 35–39	0.0096	0.0975	0	1
Remote × Age 40–44	0.0096	0.0975	0	1
Remote × Age 45–49	0.0096	0.0975	0	1
Remote × Age 50–54	0.0096	0.0975	0	1
Remote × Age 55–59	0.0096	0.0975	0	1
Remote × Age 60–64	0.0096	0.0975	0	1
Remote × Age 65–69	0.0096	0.0975	0	1
Remote × Age 70–74	0.0096	0.0975	0	1
Remote × Age 75–79	0.0096	0.0975	0	1
Remote × Age 80+	0.0096	0.0975	0	1

Appendix 2: Model coefficients and significance

Variable	Coefficient	SE
ILOC size	2.313***	0.45
Female	0.044	0.17
Growth town	– 0.685	0.91
Survival rate	8.972*	5.26
<i>Cohort age (2011)</i>		
Age 5–9	–	
Age 10–14	4.851***	1.66
Age 15–19	10.965***	2.23
Age 19–24	13.38***	2.70
Age 25–29	7.256***	2.63
Age 30–34	5.931**	2.58
Age 35–39	6.318**	2.46
Age 40–44	6.566**	2.62

Variable	Coefficient	SE
Age 45–49	4.513*	2.31
Age 50–54	5.695**	2.39
Age 55–59	4.764*	2.68
Age 60–64	1.303	2.69
Age 65–69	– 2.835	2.59
Age 70–74	– 1.526	2.95
Age 75–79	– 5.648*	3.06
Age 80+	– 5.262	3.69
<i>ILOC size by Age Interaction terms</i>		
ILOC size × Age 10–14	– 1.409***	0.32
ILOC size × Age 15–19	– 2.382***	0.39
ILOC size × Age 19–24	– 2.623***	0.47
ILOC size × Age 25–29	– 1.642***	0.46
ILOC size × Age 30–34	– 1.449***	0.45
ILOC size × Age 35–39	– 1.532***	0.43
ILOC size × Age 40–44	– 1.537***	0.46
ILOC size × Age 45–49	– 1.248***	0.40
ILOC size × Age 50–54	– 1.437***	0.42
ILOC size × Age 55–59	– 1.434***	0.47
ILOC size × Age 60–64	– 0.859*	0.46
ILOC size × Age 65–69	– 0.418	0.43
ILOC size × Age 70–74	– 0.744	0.47
ILOC size × Age 75–79	– 0.296	0.45
ILOC size × Age 80+	– 0.640	0.58

*** $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.001$

Variable	Coefficient	Standard Error
<i>ARIA</i>		
Outer regional	–	
Remote	0.062	1.03
Very remote	3.209***	1.02
<i>State/territory</i>		
New South Wales	–	
Victoria	– 0.786	0.75
Queensland	0.102	0.39
South Australia	– 0.602	0.42
Western Australia	0.266	0.36
Tasmania	– 0.587	0.42
Northern Territory	– 0.052	0.35

Variable	Coefficient	Standard Error
<i>ARIA by Age interaction terms</i>		
Outer regional × Age 10–14	6.272***	1.18
Outer regional × Age 15–19	3.621***	1.03
Outer regional × Age 19–24	1.532	1.10
Outer regional × Age 25–29	3.266***	1.01
Outer regional × Age 30–34	3.007***	0.97
Outer regional × Age 35–39	3.746***	1.04
Outer regional × Age 40–44	3.27***	1.03
Outer regional × Age 45–49	2.901***	0.97
Outer regional × Age 50–54	3.458***	1.01
Outer regional × Age 55–59	3.443***	1.14
Outer regional × Age 60–64	1.42	1.13
Outer regional × Age 65–69	1.113	1.18
Outer regional × Age 70–74	1.324	1.24
Outer regional × Age 75–79	0.878	1.33
Outer regional × Age 80+	0.128	1.56
Remote × Age 10–14	3.119***	1.08
Remote × Age 15–19	2.572**	1.10
Remote × Age 19–24	1.653	1.14
Remote × Age 25–29	2.324**	0.98
Remote × Age 30–34	2.318**	0.95
Remote × Age 35–39	3.075***	1.02
Remote × Age 40–44	3.508***	0.97
Remote × Age 45–49	2.917***	0.95
Remote × Age 50–54	2.953***	0.97
Remote × Age 55–59	2.913***	1.08
Remote × Age 60–64	1.717	1.08
Remote ×× Age 65–69	2.405**	1.12
Remote × Age 70–74	2.544**	1.19
Remote × Age 75–79	1.44	1.30
Remote × Age 80+	2.177	1.48
Constant	– 24.267***	6.2

*** $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.001$

References

- ABS (2011) Australian Statistical Geography Standard (ASGS): indigenous structure, vol 2. Australian Bureau of Statistics, cat no. 1270.0.55.002
- ABS (2018) Census of population and housing: characteristics of Aboriginal and Torres Strait Islander Australians, 2016, Explanatory Notes. <https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/2075.0Explanatory%20Notes252016?OpenDocument>. Accessed 29 June 2020
- Amemiya T (1984) Tobit models: a survey. *J Econom* 24(1):3–61

- Biddle (2009) The geography and demography of Indigenous migration: insights for policy and planning. Technical report, Centre for Aboriginal Economic Policy Research, Australian National University, Canberra
- Biddle N, Taylor J (2009) Indigenous population projections, 200631: planning for growth. Technical report, Centre for Aboriginal Economic Policy Research, Australian National University, Canberra
- Biddle N, Wilson T (2013) Indigenous Australian population projections: problems and prospects. *J Popul Res* 30(2):101–116
- Booth H (2006) Demographic forecasting: 1980 to 2005 in review. *Int J Forecast* 22:547–581
- Greene W (2003) *Econometric analysis*. Pearson Education, London
- Greene WH (2012) *Econometric modelling guide*. Econometric Software Inc, Plainview
- NT-Government (2009) A working future: real towns, real jobs, real opportunities. Technical report, Northern Territory Government
- Rabe-Hesketh S, Skrondal A (2008) *Multilevel and longitudinal modeling using Stata*, 2nd edn. Stata Press, College Station
- Raymer J, Biddle N, Campbell P (2017) Analysing and projecting Indigenous migration in Australia. *Appl Spat Anal Policy* 10:211–232
- Raymer J, Shi Y, O'Donnell J, Biddle N (2018) Multistate projections of Australia's Indigenous population: interacting area group and identification status change. *Vienna Yearb Popul Res* 16:135–162
- Regional-Services-Reform-Unit (2016) Resilient families, strong communities: a roadmap for regional and remote Aboriginal communities. Technical report, Department of Regional Development, State of Western Australia, Perth
- Sanders W (2010) Working future: a critique of policy by numbers. Technical report, Centre for Aboriginal Economic Policy Research, Australian National University, Canberra
- Steering Committee for the Review of Government Service Provision (2016) Overcoming Indigenous disadvantage: key indicators. Technical report, Productivity Commission, Canberra
- Taylor J (1997) The contemporary demography of indigenous Australians. *J Aust Popul Assoc* 14(1):77–114
- Taylor J (2014) Population projections for sparsely populated areas: reconciling error' and context'. *Int J Popul Res* 2014:9
- Taylor J, Bell M (2002) The Indigenous population of Cape York Peninsula, 2001–2016. Technical report, Centre for Aboriginal Economic Policy Research, Australian National University, Canberra
- Taylor J, Brown D, Bell M (2006) Population dynamics and demographic accounting in arid and savannah Australia: methods, issues and outcomes. Technical report, Desert Knowledge Cooperative Research Centre. Alice Springs. DKCRC Research Report 16
- Venn D, Dinku Y (2019) Can increased educational attainment explain declining Indigenous fertility? *Aust J Soc Issues*. <https://onlinelibrary.wiley.com/doi/pdf/10.1002/ajs4.89>. Accessed 29 June 2020
- Wilson T (2009) A multistate model for projecting regional populations by Indigenous status: an application to the Northern Territory, Australia. *Environ Plan* 41:230–249
- Wilson T (2011) Modelling with NEWDSS: producing state, regional and local area population projections for New South Wales. In: Stillwell J, Clarke M (eds) *Population dynamics and projection methods*. Springer, Dordrecht, pp 61–97
- Wilson T, Barnes T (2007) Continuing challenges in attempting to measure the size, and changing size, of Australia's Indigenous population. *People Place* 15(3):12–21

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.