Estimation Effects of Various Demographic Forecasting Techniques in Japan Using an Overlapping Generations Model [*]

Adam Alexander Oppenheimer [†]

March 2020

Abstract

First paragraph in the abstract
Second paragraph in the abstract

keywords: Overlapping Generations Model, Demographic Transitions, Japanese Calibration

[†]University of Chicago, Department of Economics. Email: oppenheimer@uchicago.edu.

I – INTRODUCTION

Research question: how does changing the population forecasting model alter future tax revenue through its effect on labor supply?

Literature review: primarily building off Kei and my previous paper. That paper uses as inspiration Muto et. al 2016, who carry out an overlapping generations analysis of the effects of these demographic shifts on long-run aggregate outcomes for Japan. They find that decreasing fertility rates contribute to decreasing aggregate output, while increasing longevity counteracts this by increasing aggregate output. However, this increase is not enough to create a net positive effect. Overall, their simulation predicts negative effects on aggregate outputs during the next few decades.

There are many other OLG models looking at the Japanese demographic transition but that is the most recently published paper. These papers almost entirely use the National Institute of Population and Social Security Research fertility/mortality/population forecasts. Some use UN estimates.

II – MODEL

Simpler OLG than that used in Kei and my previous paper.

III – DATA

Same data as Kei and my previous paper. Fertility data from the Human Fertility Collection. Mortality and population data from the Japanese Mortality Database. Frisch elasticity of 0.5 from Muto et. al. Wage data by age from Basic Survey on Wage Structure. Labor force participation rates by age from Labour Force Survey. Tax rates from Tax Guide Book.

IV – RESULTS

No OLG results yet. See Appendix A for demographic transition forecasts.

V – CONCLUSION

Extensions: endogenous fertility.

A − APPENDIX

Aa Demographics

Population evolution follows Aa eqn. 1 and Aa eqn. 2, where the s subscript indicates age and the t subscript indicates period. $\omega_{s,t}, f_{s,t}, \rho_{s,t} \geq 0$ represent the population, fertility rate, and mortality rate. $i_{s,t}$ represents the immigration rate. E indicates the first working year of life, while $E + S$ indicates the maximum possible lifespan in the model.

$$\omega_{1,t+1} = (1 - \rho_{0,t}) \sum_{s=1}^{E+S} f_{s,t}\omega_{s,t} + i_{1,t}\omega_{1,t} \quad \forall t \tag{Aa eqn. 1}$$

$$\omega_{s+1,t+1} = (1 - \rho_{s,t})\omega_{s,t} + i_{s+1,t}\omega_{s+1,t} \quad \forall t \quad \text{and} \quad 1 \leq s \leq E + S - 1 \tag{Aa eqn. 2}$$

This is a slight variation on the model used by **?** and **?**. The model used in both these papers assumes fertility, mortality and immigration rates are fixed at their 2014 levels. This paper attempts to forecast changes in those variables over time. The following sections discuss calibration and estimation for fertility, mortality, immigration, and population.

i Fertility

I begin the fertility rate estimation by fitting the yearly distributions using a generalized beta 2 distribution. The pdf for this distribution can be seen in Aa eqn. 3:

$$f(x|a,b,p,q) = \frac{ax^{ap-1}}{b^{ap}B(p,q)\left(1 + \left(\frac{x}{b}\right)^a\right)^{p+q}}, \quad x \in [0,\infty); a,b,p,q > 0 \tag{Aa eqn. 3}$$

where $B(v,w) = \int_0^1 t^{v-1}(1-t)^{w-1}dt$ is the beta function. Because the generalized beta 2 is a pdf but fertility rates are not a pdf, I also have to add in a scale parameter. I estimate fertility parameters from 1970 to 2014, the most recent year of data. I chose to begin in 1970 because the most recent trend in data seems to have started around then.

The fit of the model in selected years can be seen in Figure 1. The fit at the peak of the distribution is very close. The fit at the tails of the distribution is good at first but weakens over time - while it is possible to estimate parameters that better fit the tail for more recent data, the variance of the parameter estimates grows considerably and a trend in parameters disappears. Because of this, I chose to use parameters that had a slightly worse fit to the data in order to better estimate a trend.

Figure 1: Fertility estimated by generalized beta 2
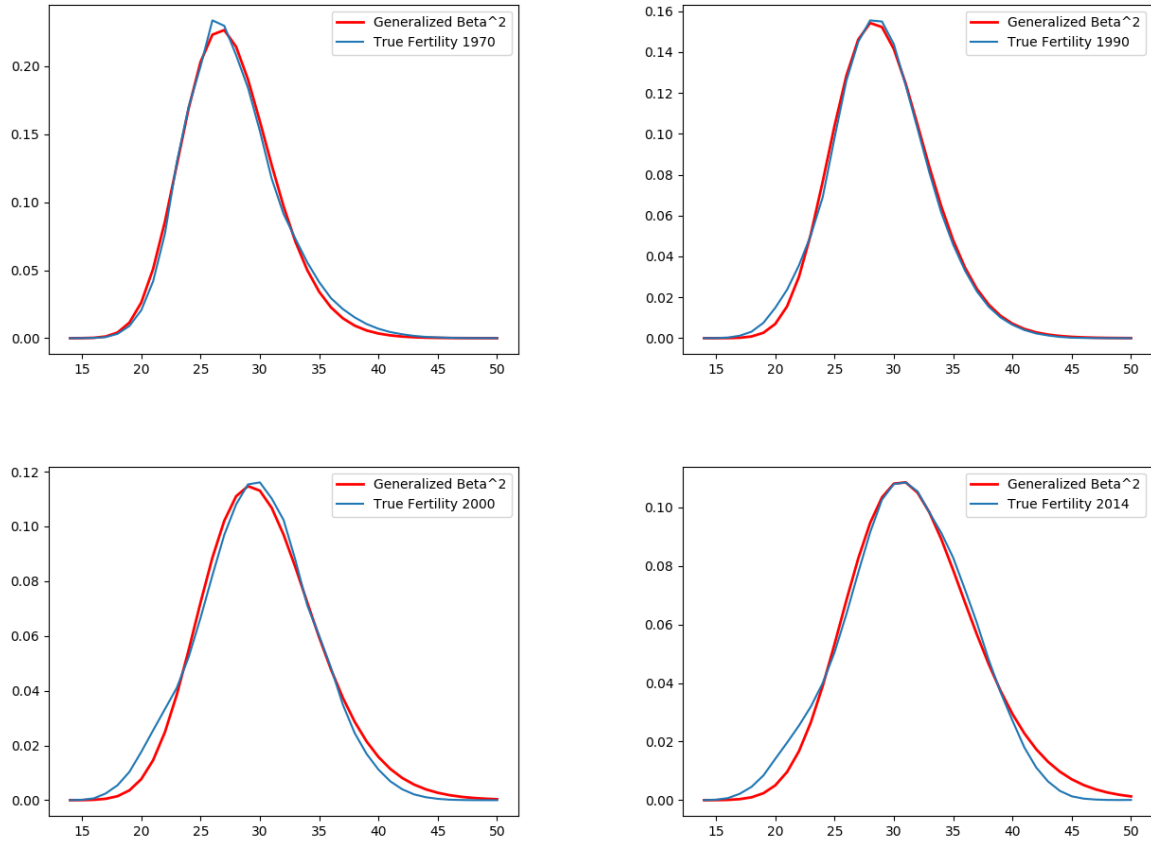
Parameter estimates and their fit over time can be seen in Figure 2. In order to ensure estimated trends converge over time, I fit these parameter estimates to logistic functions. The logistic function can be seen in Aa eqn. 4:

$$f(x|L, k, x_0) = \frac{L}{1 + e^{-k(x-x_0)}} \tag{Aa eqn. 4}$$

where L represents the maximum value of the curve, k represents the steepness of the curve, and x_0 represents the midpoint of the curve.

Figure 2: Fertility estimated generalized beta 2 parameter estimates
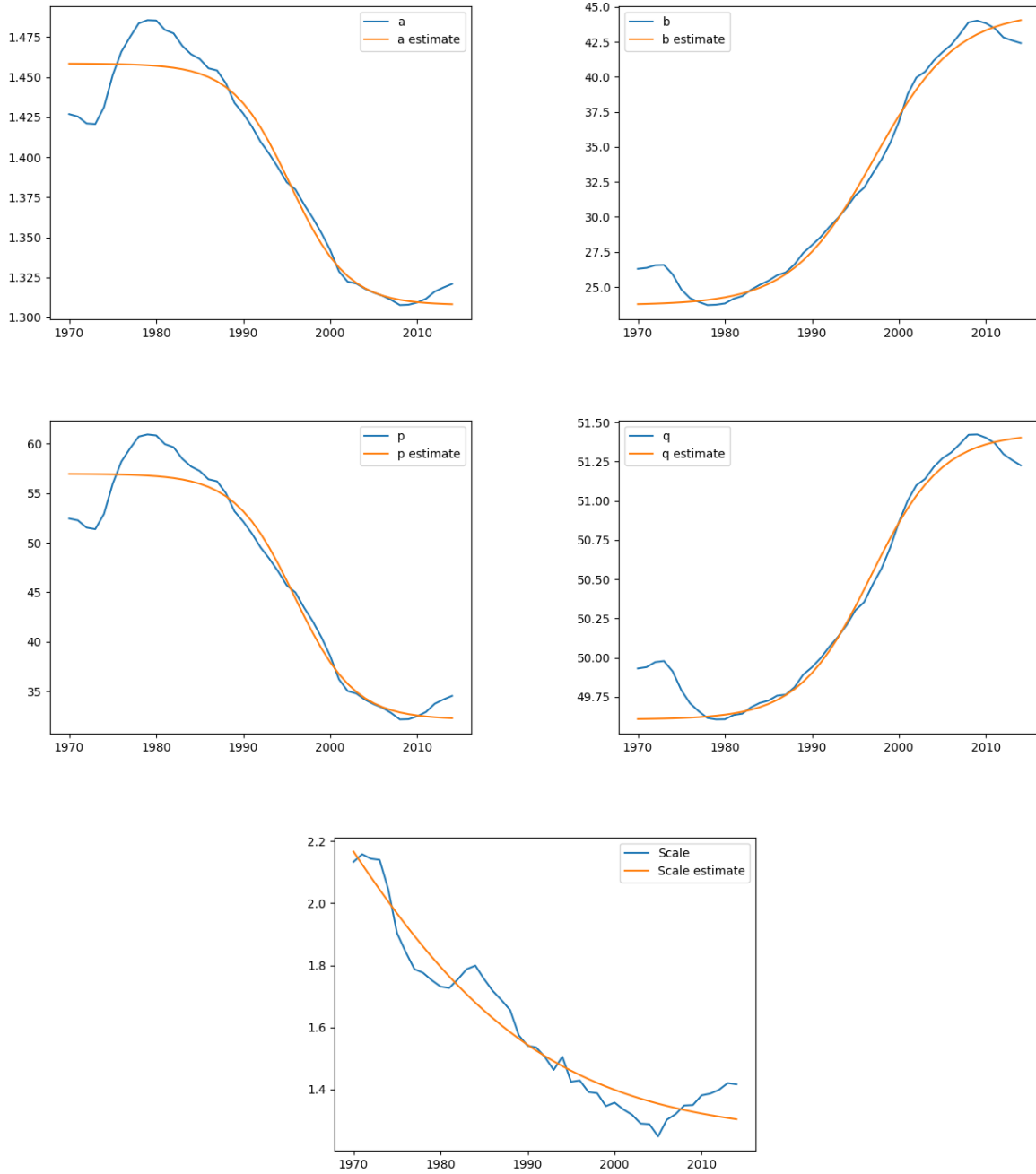
With the parameters from the logistic function estimated, we can now test the quality of our model fit. Comparisons of the models estimated with true data can be seen in Figure 3. The first panel shows the trend in true fertility rates from 1970-2014; the second panel shows the trend in model fertility rates from 1970-2014 using the logistic estimates of the generalized beta 2 parameters; the third panel shows the first two trends overlayed, with true data in red and model data in blue; and the fourth panel shows select estimates from the model.

From the first three panels, we can see that the model represents the general trends in fertility rates properly. Just as in the true data, the model correctly increases the modal fertility age over time while decreasing overall

fertility over time. As discussed before, we can see in the overlay that the model does a better job of fitting the center of the distribution than the tails of the distribution.

The fourth panel gives us an indication of how fertility will evolve over time. We can see the large drop in fertility from 1990 to 2000 to 2014, which follows the same trend as the true data. With our model we now also have the ability to forecast future fertility rates. Our model forecasts that by 2050 fertility rates will continue their trend downward with a concurrent increase in the modal fertility age. We also see that the logistic functional form leads to a convergence in fertility rates by 2050 - the estimate for 2100 is almost exactly overlaid with the estimate for 2050.

Figure 3: Fertility estimated generalized beta 2 model fit
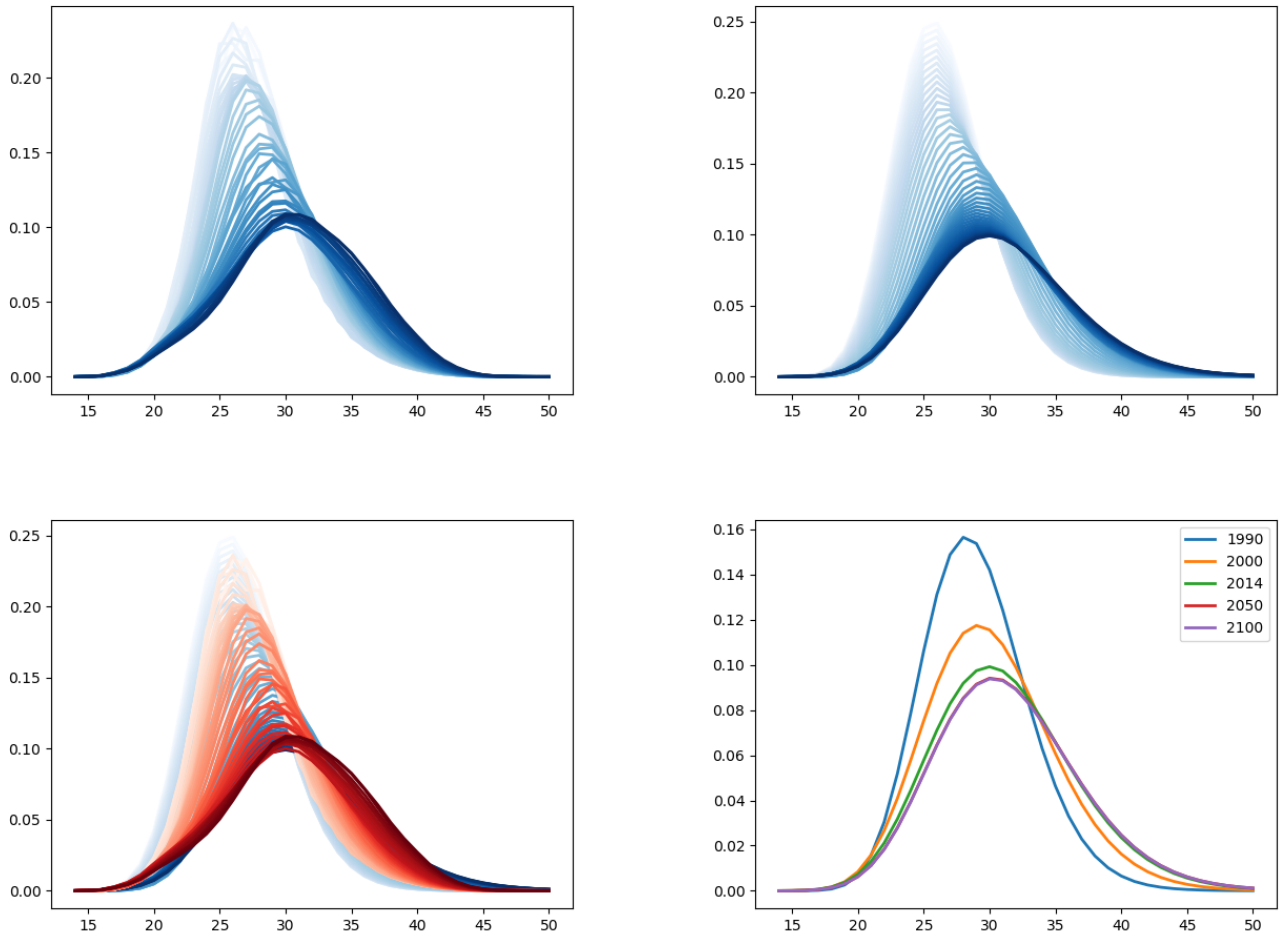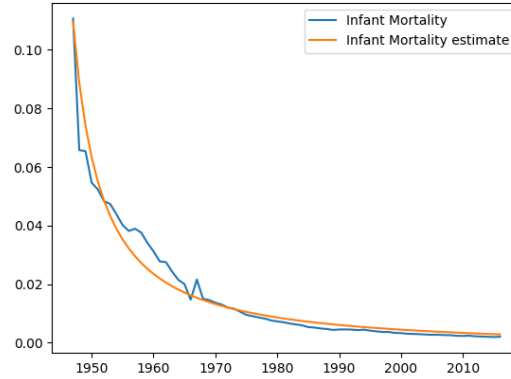

ii Mortality

Mortality rate estimation requires working in two steps: first, estimating infant mortality, then estimating non-infant mortality. This is necessary because infant mortality rates over the past century have been decreasing at a much faster rate than mortality rates for other ages. It is therefore necessary to model them separately in order to accurately represent the varying trends.

The infant mortality rate estimation takes one step: I fit the trend to a generalized polynomial of the form seen in Aa eqn. 5:

$$f(x|a,b,c,d,e) = a(e \cdot x - b)^{\frac{1}{c}} + d \qquad \text{(Aa eqn. 5)}$$

The fit of the model can be seen in Figure 4. Because of the steep drop in infant mortality rates over time, I chose to estimate it for the entire set of data from 1947-2016. However, to emphasize the importance of fitting recent data, I add additional weight to data from the last 15 years of the sample when estimating parameters. I also penalize parameter estimates that forecast negative infant mortality rates by 2100 to ensure only positive future estimates before I assume convergence by 2050.

Figure 4: Infant mortality estimated by polynomial

I begin the non-infant mortality rate estimation by fitting the yearly distributions using a generalized beta 2 distribution. The pdf for this distribution can be seen in Aa eqn. 3. This is the same distribution used to fit fertility estimates. I choose to use a generalized beta 2 distribution rather than an exponential, which is what the data appears to follow, because the fit is almost identical but there is no discernable trend in parameter estimates for an exponential distribution but there is a very clear trend for the generalized beta 2 distribution.

As with fertility rates, because the generalized beta 2 is a pdf but mortality rates are not a pdf, I also have to add in a scale parameter. I estimate mortality parameters from 1970 to 2014, the same years as for fertility rates.

The fit of the model in selected years can be seen in Figure 5. The estimated distribution has a very close fit to the true data. This is true for all years, unlike the fertility estimates which do not fit the tails properly for recent data.

Parameter estimates and their fit over time can be seen in Figure 6. In order to ensure estimated trends converge over time, I fit these parameter estimates to logistic functions. The logistic function can be seen in Aa eqn. 4. This is the same distribution used to fit fertility parameters.

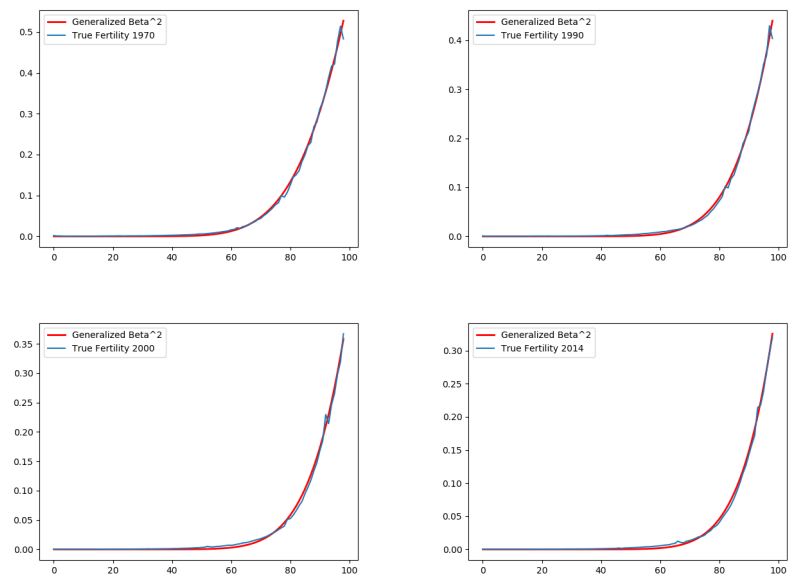Figure 5: Mortality estimated by generalized beta 2
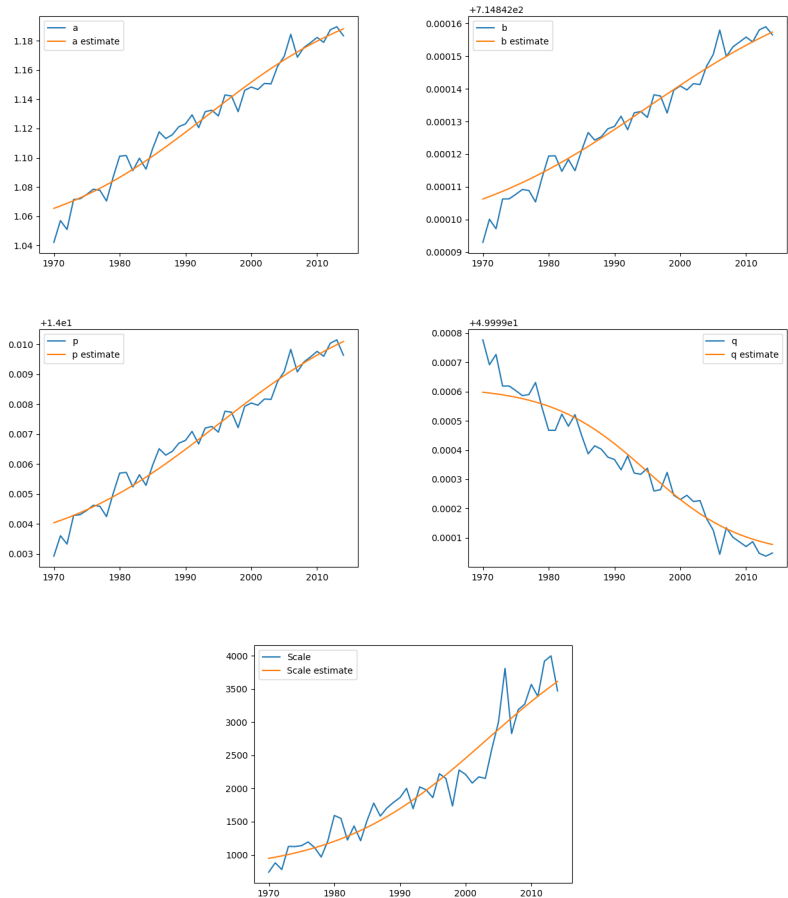

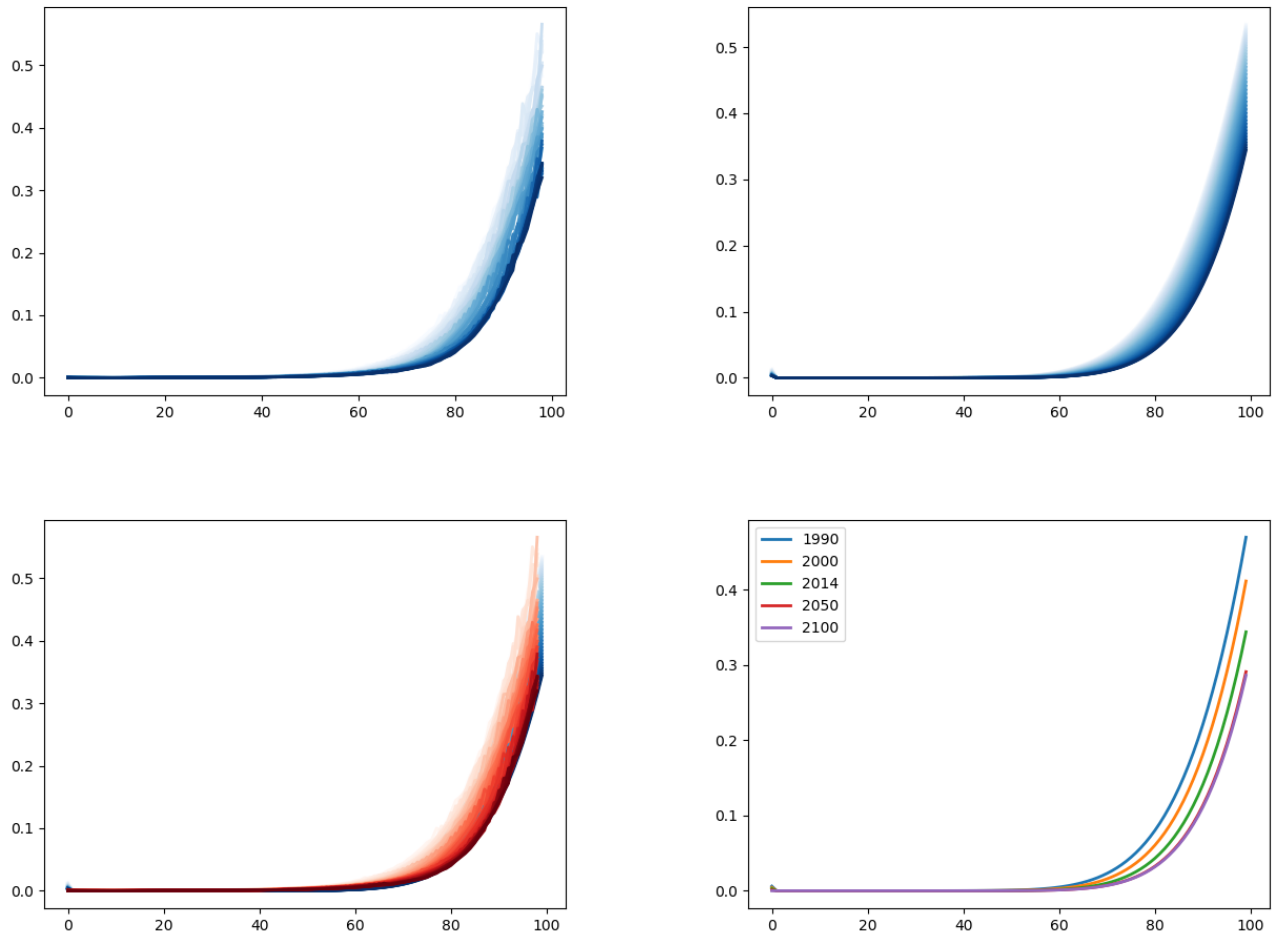Figure 6: Mortality estimated generalized beta 2 parameter estimates

With the parameters from the logistic function estimated, we can now test the quality of our model fit. Comparisons of the models estimated with true data can be seen in Figure 7. The first panel shows the trend in true mortality rates from 1970-2014; the second panel shows the trend in model mortality rates from 1970-2014 using the logistic estimates of the generalized beta 2 parameters; the third panel shows the first two trends overlayed, with true data in red and model data in blue; and the fourth panel shows select estimates from the model.

From the first three panels, we can see that the model represents the general trends in mortality rates properly.

The fourth panel gives us an indication of how mortality will evolve over time. We can see consistent decline in mortality from 1990 to 2000 to 2014, which follows the same trend as the true data. With our model we now also have the ability to forecast future mortality rates. Our model forecasts that by 2050 mortality rates will continue their trend downward. We also see that the logistic functional form leads to a convergence in mortality rates by 2050 - the estimate for 2100 is almost exactly overlaid with the estimate for 2050.

Figure 7: Mortality estimated generalized beta 2 model fit

iii Immigration

Estimating the evolution of immigration rates over time poses more of a challenge than fertility or mortality rates. While fertility and mortality rates can be estimated using a generalized beta 2 distribution, immigration rates following a distribution that is difficult to model without a mixture method. To get around this, I instead model the

evolution in immigration rates for each age. While this no longer allows for the use of a simple distribution and clear parameter evolutions over time for the entire set of data, it seems like a more reasonable way to fit the data than to fit a mixture model with a large number of distributions.

To fit the evolution of immigration rates for each age, I estimate the trend from 1997-2014 using linear regression, then forecast using an exponential distribution that matches the slope of the linear regression in 2014 and is assumed to plateau in 15 years at a value 10% beyond the last value in the data. The exponential I fit can be seen in Aa eqn. 6:

$$f(x|a,b,c,d,p,s,\beta_0,\beta_1) = e^{a(x-s)^2+b(x-s)+c} + p \qquad \text{(Aa eqn. 6)}$$

$$s.t. \quad \left.\frac{\partial f}{\partial x}\right|_{x=s} = \beta_1 \qquad \text{(cond. 1)}$$

$$f(s) = \beta_0 + \beta_1 s \qquad \text{(cond. 2)}$$

$$\left.\frac{\partial f}{\partial x}\right|_{x=s+15} \approx 0 \qquad \text{(cond. 3)}$$

$$f(s+15) = d \cdot f(s) \qquad \text{(cond. 4)}$$

where a, b, and c are parameters to estimate, p shifts the curve so we can estimate c (this is explained in the derivation), s gives the last year of data, β_0 and β_1 are the OLS estimates from fitting the data, and we define d such that

$$d = \begin{cases} 0.9 & \text{if} \quad (\beta < 0 \text{ and } f(s) > 0) \text{ or } (\beta > 0 \text{ and } f(s) < 0) \\ 1.1 & \text{if} \quad (\beta < 0 \text{ and } f(s) < 0) \text{ or } (\beta > 0 \text{ and } f(s) > 0) \\ 1 & \text{if} \qquad\qquad\qquad \beta = 0 \end{cases}$$

For notational simplicity, I will define $\beta = \beta_0 + \beta_1 s$. We therefore have from cond. 2 that $e^c + p = \beta \Leftrightarrow c = \log(\beta - p)$. In order to ensure that we can compute c, we set $p < \beta$.

Using $\frac{\partial f}{\partial x} = (2a(x-s) + b)e^{a(x-s)^2+b(x-s)+c}$ and cond. 1, we have $be^c = \beta_1 \Leftrightarrow b = \frac{\beta_1}{\beta-p}$.

From cond. 3 we have $(2a(15) + b)(f(x + 15) - p) \approx 0$. Plugging in the value for $f(x + 15)$ from cond. 4, and recalling that $f(s) = \beta_0 + \beta_1 s = \beta$, we have $\left(2a(15) + \frac{\beta_1}{\beta-p}\right)(d \cdot \beta - g) \approx 0 \Leftrightarrow a \approx \frac{1}{30}\left(\frac{0}{d \cdot \beta - g} - \frac{\beta_1}{\beta-p}\right)$. Because the derivative cannot actually become 0, we can choose a value very close to 0 to estimate this curve numerically. At 2030, we assume that the estimate stays constant for future years.

We therefore have the following:

$$\left\{ p < \beta, a \approx \frac{1}{30}\left(\frac{0}{d \cdot \beta - g} - \frac{\beta_1}{\beta - p}\right), b = \frac{\beta_1}{\beta - p}, c = \log(\beta - p) \right\}$$

where $\beta = \beta_0 + \beta_1 s$.

The fit of the model for selected ages can be seen in Figure 8. Unlike fertility and mortality rates, fertility rates by age appear not to have as predictable a trend. However, the linear regression model with exponential forecasts seems sufficient to have reasonable forecasts for the short term.

Figure 8: Immigration estimated by linear regression and forecasted by exponential
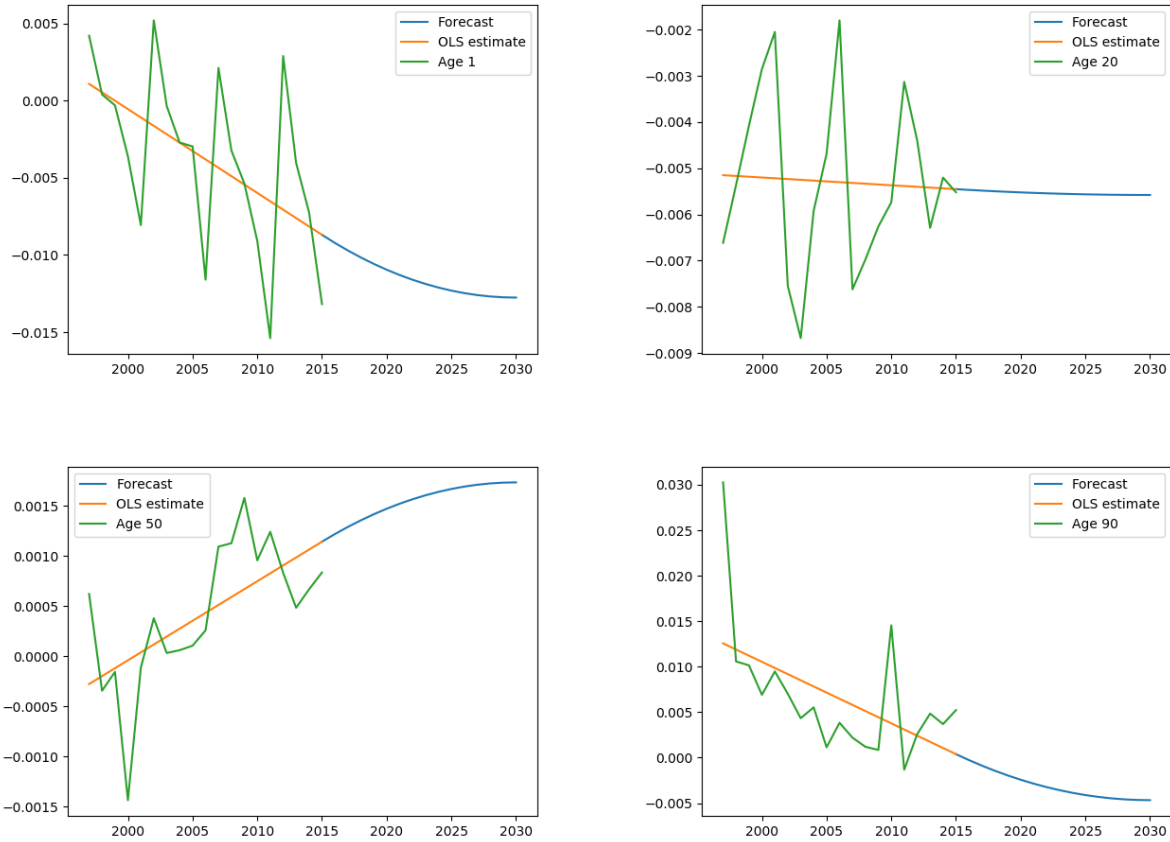

With the parameters from the linear regression and exponential estimated, we can now test the quality of our model fit. Comparisons of the models estimated with true data can be seen in Figure 9. The first panel shows the trend in true immigration rates from 1997-2014; the second panel shows the trend in model immigration rates from 1997-2014; the third panel shows the first two trends overlayed, with true data in red and model data in blue; and the fourth panel shows select estimates from the model.

From the first three panels, we can see that the model represents the general trends in immigration rates properly, although with reduced variance.

The fourth panel gives us an indication of how immigration will evolve over time. We can see consistent decline in immigration from 1990 to 2000 to 2014 for the young and the old, with little change in immigration rates for other ages, which follows the same trend as the true data. With our model we now also have the ability to forecast future immigration rates. Our model forecasts that by 2050 immigration rates will continue their trend downward for the old. In order to prevent population estimates from declining too rapidly I choose to increase immigration rates for the 0 year old population rather than let it continue to decline. Other ages seem to have relatively unchanged immigration rates over time. We can also see that following our construction, we have a convergence in immigration rates by 2030.

Figure 9: Immigration estimated by linear regression and forecasted by exponential
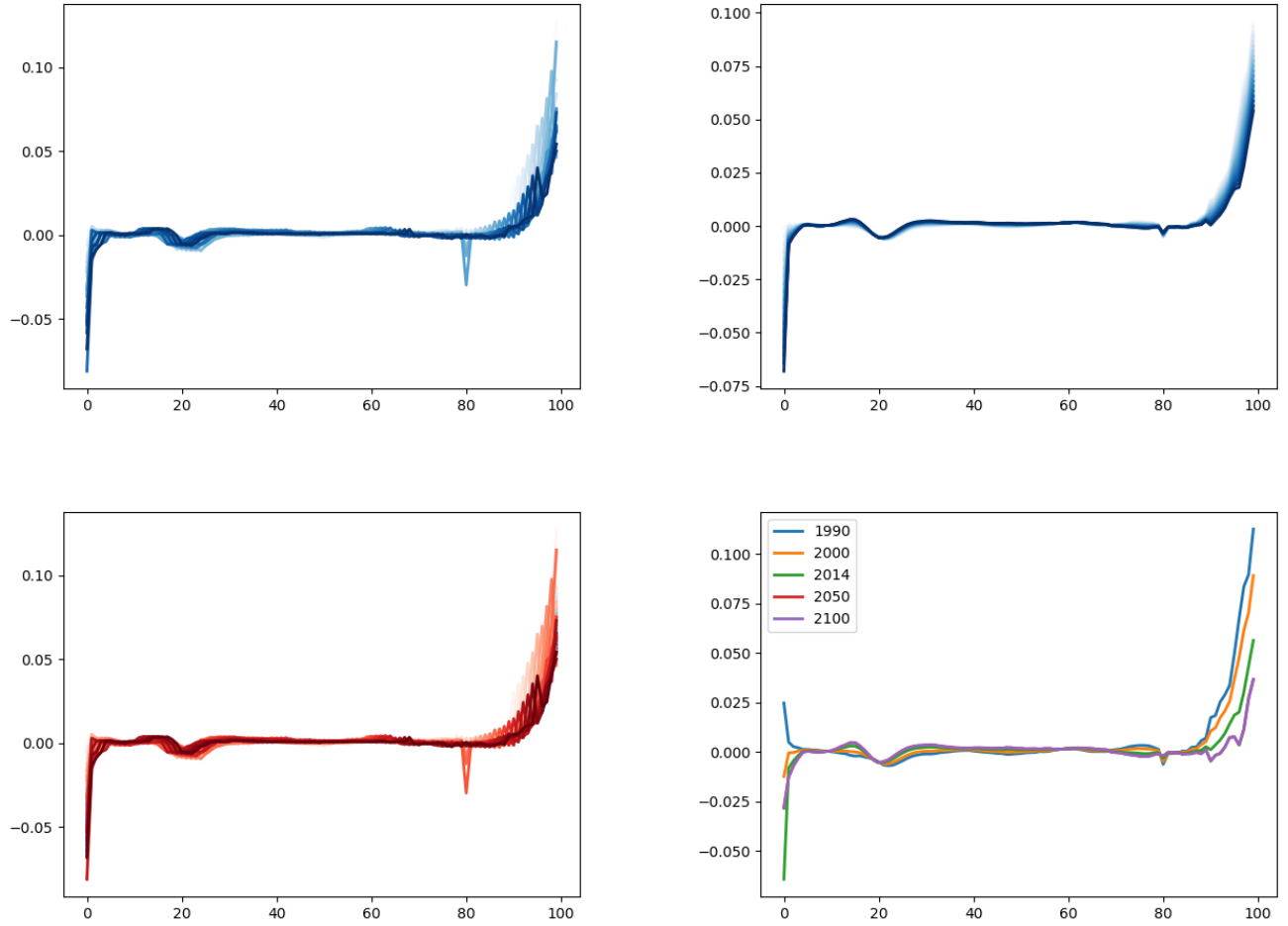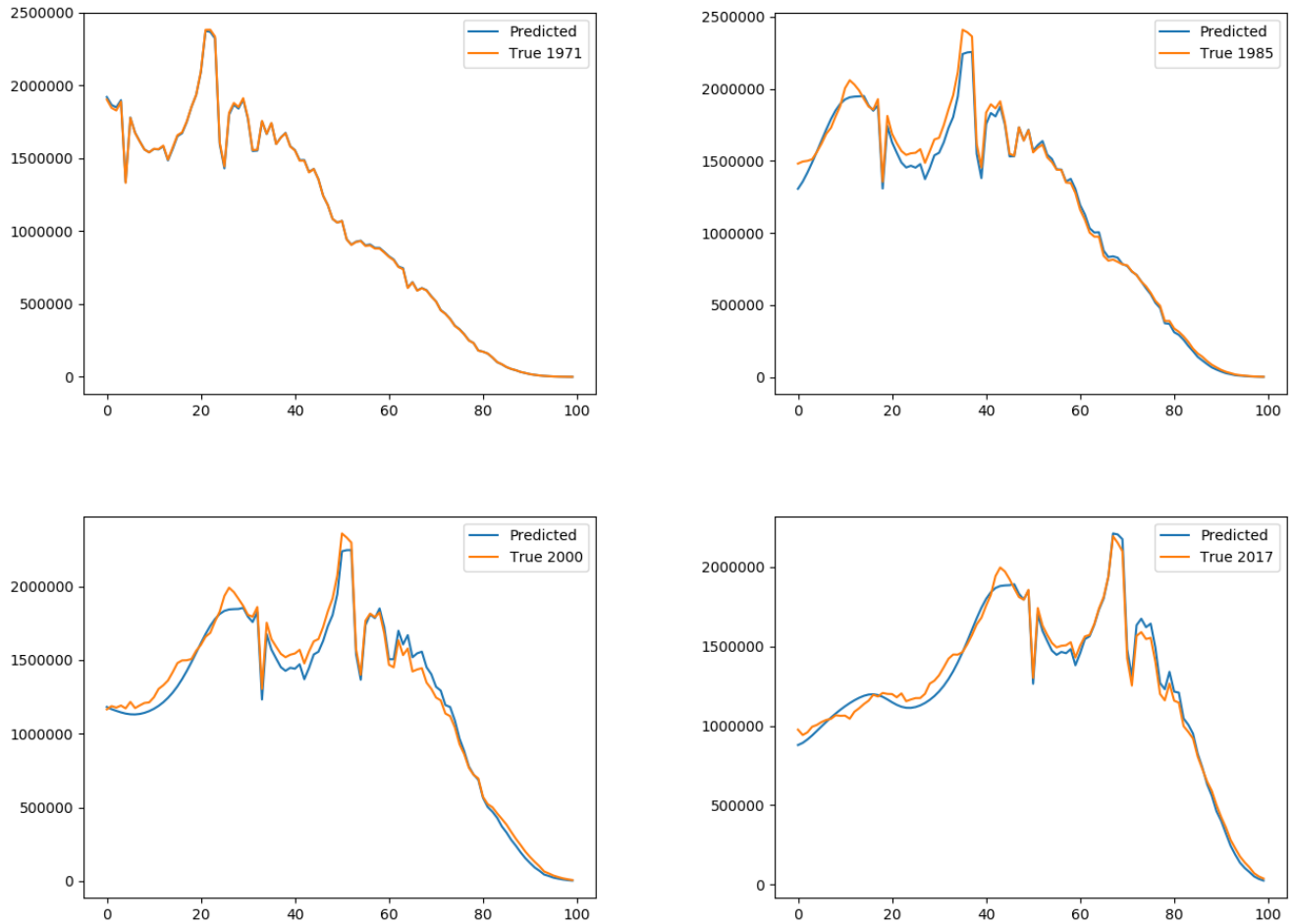
iv Population

Using our fertility, mortality, and immigration estimates, we can now use the population evolution described in Aa eqn. 1 and Aa eqn. 2 to forecast future population given an initial year of population data.

We will first analyze quality of fit on past data. I test by beginning with 1970 population data and comparing the true population evolution against the population evolution forecasted by the model. Selected years of the transition from 1970 to 2017, the most recent population data, can be seen in Figure 10. We can see in the first year of the model that there is a slight overestimate of births. This persists over time for this particular cohort, as the mortality rate should not be compensating for this. However, this slight overestimate is actually a rise in population that is is seen to occur just a few years later in the true data, when looking at the figure from 1985. The model does an accurate job of representing the inflection point in population that occurs around 1974 in the real data, and 1972 in the forecast. Looking at the figure from 2000, we can see that the model also accurately fits the inflection point in population around 1997. However, we begin to see that there a few ages where the model is overestimating the population and other ages where it is underestimating the population. The final figure is for 2017. This is 47 years after the start. The model appears to fit the data well and reflects the inflection points in the data.

Figure 10: Estimated population

Given the model does a good job of fitting historical data, we now move on to forecasting future populations. We can see how population will evolve over time in Figure 11. The first panel shows the population at select years in the future. These results are rather surprising. We can see that given the forecasts for fertility, mortality, and immigration rates, the population in Japan is not sustainable. According to this model, by the year 2500 the population will have effectively disappeared. This result should caution the user of this model - it can lead to counterintuitive and unreasonable results. However, without another model to use in its place, this model is used for the estimation in this paper.

The second panel shows the distribution of the population at select years in the future. We can see that over time, the modal age of the population increases until it peaks at around 75 years old. While this result indicates Japan faces a massive hurdle with its aging population, it should be taken with a grain of salt considering that the model estimates that by the time the demographic distribution reaches this point, the population will be effectively 0.

Figure 11: Forecasted population