IIMB-WP N0. 656/2022

WORKING PAPER NO: 656

Forecasting Mortality Rates for India Using Lee-Carter Model with applications to Projecting Population Pyramid, Gender Imbalance and Dependency Ratio

Shubhabrata Das
Professor of Decision Sciences
Indian Institute of Management Bangalore
Bannerghatta Road, Bangalore – 5600 76
shubho@iimb.ac.in

Sidhartha S Guru
Senior Product Manager,
Booking.com B. V
guru.sidhartha@gmail.com

Abhiram Iyer
Assistant Vice President
Deutsche CIB Centre Pvt. Ltd
iyer.abhiram@gmail.com

Year of Publication – February 2022

Forecasting Mortality Rates for India Using Lee-Carter Model with applications to Projecting Population Pyramid, Gender Imbalance and Dependency Ratio

Shubhabrata Das* Sidhartha S Guru Abhiram Iyer

Abstract

While the Lee-Carter model is considered to be the leading framework for computing mortality risk in the past two decades, it is yet to be used in modeling Indian mortality prior to the present study. The LC model is applied for the total Indian population, and also separately male, female, rural and urban sub-populations and the mortality rates are forecasted for the remainder of the century. This allows computing projected expected life at birth and comparing it across different sub-populations.

Application of the Lee-Carter model for the male and female subgroups, in combination with the forecasted birth rate and possible considerations of the gender ratio at birth, lead to the projection of the overall age-wise population pyramid for the forecast period. We propose a measure of gender-imbalance that takes into account the gender diversity across the different age-groups. This gender-imbalance index is forecasted using the projected mortality rates. The projected age-dependency ratio is a significant outcome of the undertaken research.

Keywords: ARIMA; Birth rate; Demographic dividend; Gender diversity; Gender ratio; Rural; Urban.

1 Introduction

One of the most vital resources that a country possesses is its population. It is only through leveraging this resource that any country can achieve fast and sustainable economic growth. We have witnessed challenges in the old world countries in Europe, Japan etc. due to the aging nature of their populace. Conversely, the developing countries seem to achieve higher growth rates due to their relatively younger population. A favorable demographic dividend results in the addition of more workforce and therefore sustains growth. Mason (2006) shows that demographic dividend alone explains 42% of growth in per capita GDP. India, one of the fastest growing countries in the world today, stands to gain much from the demographic dividend. The 2011 shows that India's population stands at 1.21 billion, growing at 17% over the last decade (see Indian Labour Statistic [14]). The country is uniquely placed in a position to enjoy the effects of two dividends. First, more and more youth (less than 15 years of age) enter the population, thus increasing the workforce. The workforce shrinkage also reduces over time due to reduced mortality. Secondly, elderly population (60 years+) increases due to improved life expectancy. This population group can give a sustainable boost to the economic development through careful planning of savings. The demographics can be studied by the policy-makers to accelerate the economic progress made by the country; and ensure that a demand-supply mismatch does not occur in the labor market. A reliably forecasted mortality rate and birth rate can make an objective assessment to such advantages, in addition to providing other useful insights for planning.

Since the seminal publication Lee & Carter (1992) which dealt with the mortality data in USA, the Lee-Carter (LC) model has been extensively used in forecasting mortality rates in various countries, mostly

*Correspondence to: Shubhabrata Das, Faculty Block C 2nd Floor, Indian Institute of Management Bangalore, Bannerghatta Road, Bangalore 560076 India. E-mail: shubho@iimb.ac.in

in Europe. To name a few, Brouhns et al. (2002) applied it for Belgium, Cairns et al. (2009) applied it for England and Wales, Wang (2007) applied it for Sweden, Yasungnoen and Sattayatham (2016) used the model for Thailand. Booth & Tickle (2008) reviewed various mortality models, including variations of the LC model. However, the LC model has not been applied on Indian mortality data, possibly because of relatively short history of reasonably reliable age-specific mortality figures. The primary contribution of the present work lies with the first implementation of LC model in the Indian context and this is based on mortality data from 1997 to 2013.

A forecast of mortality rates serves many purposes other than routine actuarial applications in cost-based pricing of insurance products. Since we implement the LC model separately for male, female, rural, and urban (Indian) population, we are able to forecast mortality rates separately for these subgroups, and this leads to several useful demographic projections which is focussed in terms of application in this work. One such important application is that of constructing age-wise population pyramids; this is taken up for the coming century, using the projected birth rate and gender ratio at birth.

We also focus on age-specific gender-imbalance through established and new measures proposed in this work. In stead of just looking at aggregate count of male and female population, the proposed index captures the mismatch in each age-group. The treatise facilitates projection of gender-equity in years to come through a forecast of this gender imbalance index.

The dependency ratio is defined as the ratio of under 15 and above 64 year old populations to the population aged between 15–64 and it is intended to capture the percentage of the population dependent on population in the working age. This dependency ratio is a standard index to capture the impact rate of ageing in a country and the projected values for the coming century is another significant outcome of the undertaken research.

In the next section, we discuss the Lee-Carter (LC) Model and its adaptation to the Indian mortality data. In Section 3.1, we report the estimated parameters of the LC model and reflect on the comparative values. In Section 3.2, we discuss the fitted and forecasted mortality rates for the entire Indian population. Mortality rates are useful in calculating expected life at birth; the projected values of this index for the different sub-populations are taken up in Section 3.4. Comparison of the forecasted mortality rates for the urban and rural settlements and for sub-populations based on gender is taken up in Section 3.3 and is key to building the projected population pyramid. Three other key demographic factors – birth rate, gender ratio at birth and net migration rate – can affect population projection gender diversity and dependency ratio; they are discussed in Section 4.1. The forecasted population pyramids are shown in Section 4.2. In this paper, we propose a measure of gender-imbalance that takes into account the diversity across the age-groups. This index is introduced and the projected values are presented in Section 4.3. The projected dependency ratio is shared in Section 4.4. We conclude with a summary and few related observations in Section 5.

2 The Lee-Carter Model and applying it for India

2.1 The Lee-Carter (LC) Model

The Lee-Carter (LC) model, envisages the age-specific and time dependent mortality rate as linear function of time-specific index, with the parameters dependent solely on age. According to the model:

$$\log(m_{x,t}) = a_x + b_x k_t + \epsilon_{x,t}, \tag{1}$$

where $m_{x,t}$ is the mortality rate for the age group x at time t; a_x is the average mortality pattern (age-specific); b_x is the deviation of mortality pattern over time (age-specific); k_t is the time-specific mortality index; and $\epsilon_{x,t}$ is the error term. The identifiability (uniqueness) of the parameters are ensured with additional constraints:

$$\sum_x b_x^2 = 1; \quad \sum_t k_t = 0. \tag{2}$$

It is common to enforce $\sum_x b_x = 1$ in stead of the first constraint; we have considered the modified form as unlike many researchers, we would allow some of the b_x's to be negative, as explained subsequently.

In the first stage of estimation, the average age-specific mortality rate a_x is estimated as:

$$\hat{a}_x = \frac{1}{T} \sum \log(m_{x,t}). \tag{3}$$

This estimate is time-neutral, and it leads to computing an intermediate matrix $Z = ((z_{x,t}))$, where

$$z_{x,t} = \log(m_{x,t}) - \hat{a}_x.$$

A singular value decomposition (SVD) of the matrix Z gives $Z = UdV'$, leading to estimation of b_x and k_t from the largest eigenvalue, as follows:

$$\hat{b}_x = U_{x1} \quad \text{and} \quad \hat{k}_t = d_1 V_{t1}. \tag{4}$$

Subsequently, a random walk model with drift is applied on $\hat{k}_t$, and the forecasted values along with estimated a_x and b_x, from (3) and (4), are used to project the mortality using (1). While Lee & Carter (1992) and most researchers on LC models used the random walk model with drift for $k(t)$, in this work we propose to adopt more inclusive Box-Jenkins method, i.e. consider a suitable Autoregressive Integrated Moving Average (ARIMA) model to estimate and forecast the time-specific mortality index k_t. Such generalizations have been carried out also by DÀmato et al. (2011), Gaille (2012), among others. Let L be the lag operator. Then the non-seasonal ARIMA (p, d, q) model with drift η is given by:

$$\big(I - \sum_{i=0}^{p} \phi_i L^i\big)(I - L)^d k_t = \eta + \big(I - \sum_{j=0}^{q} \theta_j L^j\big)\epsilon_t,$$

Where d is the degree of differencing, p is the order of the autoregressive model, q is the order of the moving-average model, while ϵ_t is a white-noise. $\hat{k}_t$, obtained in (4) is used as input time-series data, and standard ARIMA procedure (see, e.g. Makridakis et al. 2005) is adopted leading to a suitable model selection within the class using Akaike Information Criteria (AIC).

2.2 Mortality data in India and Implementation of LC model

Prior to this work, the LC model has not been used with Indian mortality data. Indeed, irrespective of method used, very little research has been published in this direction with Indian context, possibly due to the limited data availability on age specific death rates, and concerns over quality of data. Only in relatively recent years, has the data been systematically collected and reported. Dubey (2014) fitted Indian mortality rates based on only population census data, which is collected at an interval of 10 years, with the growth rate in the inter-census period being computed using exponential model and population interpolated for intermediate years. Ghosh & Sriram (2005) discussed adjusting Indian mortality rates based on gender and smoking habit, although it was not on the basis of any parametric model. A UN report (2015) ([17]) projects world population demographics for various countries, including India. While the methodology adopted in [17] is different from the present work, several key demographics are considered in both and with similar forecast horizon. Consequently the projected figures reported in [17] are of interest in terms of drawing comparison with the present work.

In the current research, data is used from the Sample Registration System (SRS) issued by the Office of the Registrar General, India. The SRS uses a dual sampling system, with continuous direct enumeration of deaths in a sample of villages and urban blocks and a half-yearly verification of the data. This ensures a rather accurate representation of the country's demographics. The building-block data is the age-specific-death-rates (ASDR), i.e. the ratio of actual deaths to the exposure to death. This ASDR is reported, grouped by five-year block age groups, e.g., 0–4, 5–9, . . ., 80–84, and 85+; and this ASDR data is available for the entire population only since 1986 (till 2013). Years prior to this are either not comprehensive or not divided into five-year age groups. However the ASDR data, based on gender-based segregation and domicile (urban vis-a-vis rural) segregation is available only from 1997. Since these comparative analysis is critical to the present work, all the analysis here are presented for aforementioned 17 year period 1997–2013, 18 age groups, for the five population (sub-)groups (total, urban, rural, male, and female).

As communicated in the previous subsection, a more exhaustive search within Box-Jenkins models was adopted in the final stage of modeling k_t. This led to the same random walk model with drift, i.e. an ARIMA (0,1,0) model with a drift for the total Indian population, as well as Indian male population. But, an ARIMA (1,1,0) with drift is picked for the Indian rural population, while picking an ARIMA (1,1,1) with a drift model for the Indian urban population as well as the Indian female population.

In the following table, we report the mean absolute percentage error (MAPE) to show how LC model was able to fit Indian context for the five (sub-)populations. The table shows reasonable fit for all the five cases in that regard. Table 1 also reports the percentage of variability in the mortality rates explained [as intended in Lee & Carter (1992)]; these are quite high, not surprisingly, as the number of years on which the model is built is very much on the lower side.

Table 1: LC model fit/accuracy – Indian Mortality data and different subgroups

Indian (sub-) population	% variability explained	Mean Absolute % Error (MAPE)
Total	99.43%	4.63%
Male	98.66%	5.14%
Female	99.11%	5.45%
Urban	99.29%	6.27%
Rural	99.11%	5.58%

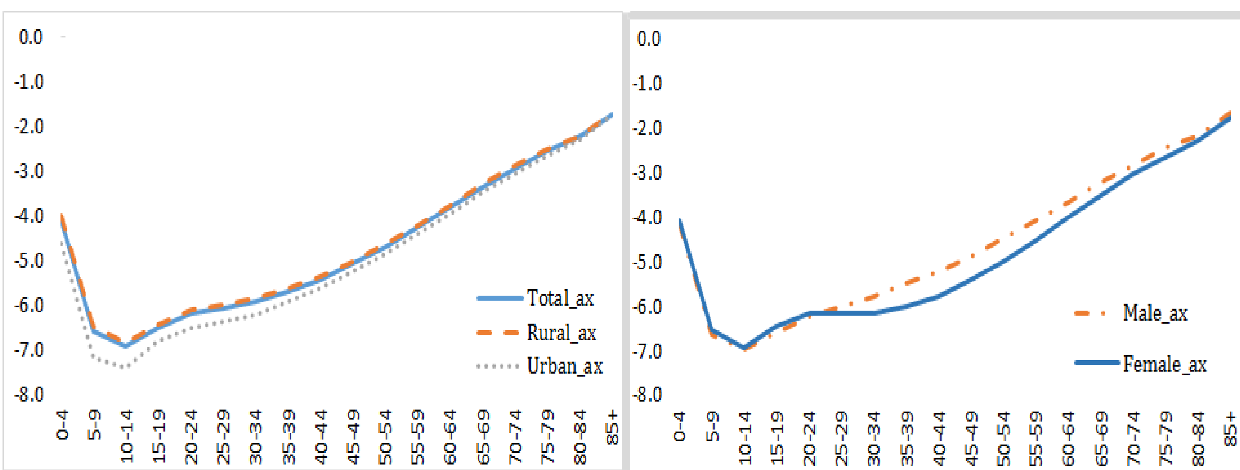
The detailed parameter estimates and their comparison across the (sub-)populations are presented in the next section.

3 Results: Forecasted Mortality Rates in India

3.1 Parameter Estimates

The estimated average age-specific mortality (a_x) is shown via Figure 1. This shows a very similar pattern for all the five (sub-)population – a sharp decrease from the first age-group, followed by a marginal drop, and then a steady increase. The overall mortality is consistently lower in urban population as compared to the rural, with the most difference being observed for the age-group 5–9 and 10–14, with gap reducing steadily with increase in age. The male and female mortality are almost identical till 20–24; but for subsequent age-groups, the average female mortalities are consistently lower than the male.

Figure 1: Estimated a_x values: Indian mortality data 1997-2013

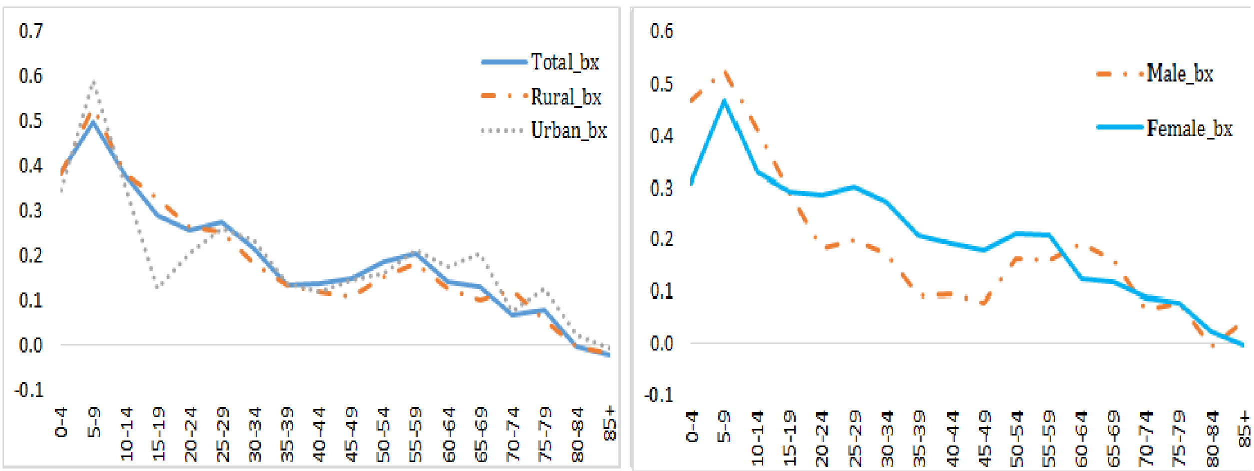


The estimated values of b_x, the age-specific deviation of mortality pattern over time, is shown via Figure 2. Recall that

$$\frac{\partial \log m(x,t)}{\partial t} = b_x \times \frac{\partial k_t}{\partial t}.$$

So a relatively high b_x for any age-group indicates that the mortality rate declines more rapidly over time for this age-group. The highest b_x is for 5–9 age bracket for all the (sub-) population; otherwise, it typically decreases with increase in age. This shows the benefit of possible scientific advancements are the highest at the younger ages, with the benefits being progressively less for the higher ages. Estimated b_x is typically lower in rural, as compared to the urban, for most age-brackets except 15–24, 0–4 and 70–74, with the difference being substantial for 15–24. Indeed, for the age-group 15–19, b_x is seen to be strikingly lower for the urban population. This could be due to increasing high suicidal tendency due to career anxiety and mental stability in the urban youth, or because of effect of migrant young workers in the urban area (from the rural). The gender-gap in b_x can be separated into three categories; between 15 to 60, this is lower for the females, while the situation is reverse at the younger ages as well as older ages. This possibly supports the societal and financial bias in favour of male, while the natural causes continue to be more favourable to women.

Figure 2: Estimated b_x values: Indian mortality data 1997-2013



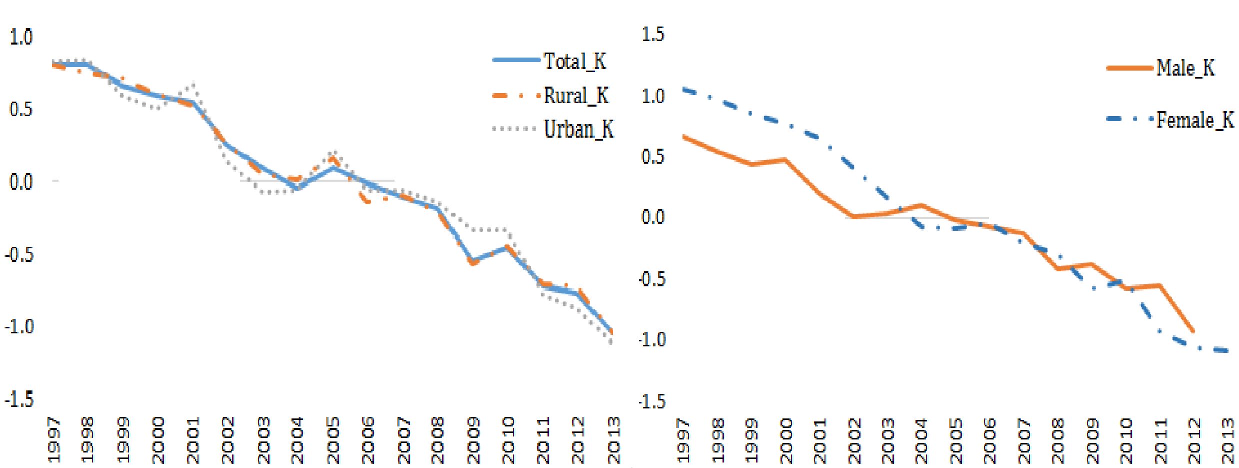
Negative value of b_x is of particular significance, as this indicates an increase in mortality rates for the corresponding age. We observe small negative values for the two highest age-brackets in some of the (sub-)population; this is reported in Table 2. Thus we observe that older age groups suffer from increasing mortality over time, especially in the rural population. We speculate that this could be due to lower quality of current healthcare, especially for the elderly in the rural sections. Also it is possible that as mortality falls for most of the (younger) age-groups, it may have to be compensated at the highest age-group.

Table 2: Estimated b_x values: Old ages

Age-group	Total	Urban	Rural	Male	Female
80 – 84	-0.0002	0.0240	-0.0013	-0.0078	0.0223
85 and above	-0.0216	-0.0050	-0.0156	0.0435	-0.0032

The fitted values of k_t, the time-specific mortality index, is represented through Figure 3. This shows a steady decrease in mortality index over time horizon 1997–2013. The gender-divide on this front is quite sharp, with the slope of the time evolution, k_t, being distinctly higher for female as compared to the male. The domicile, i.e. the urban-rural difference on this front is marginal.

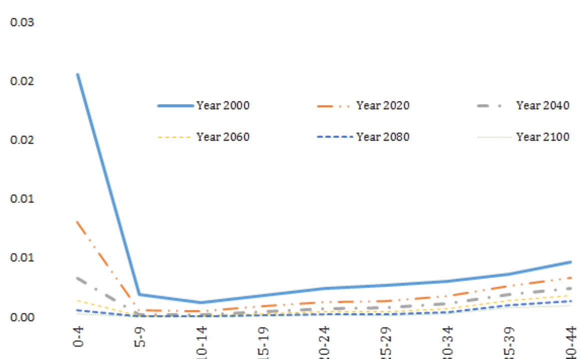
Figure 3: Fitted k_t values



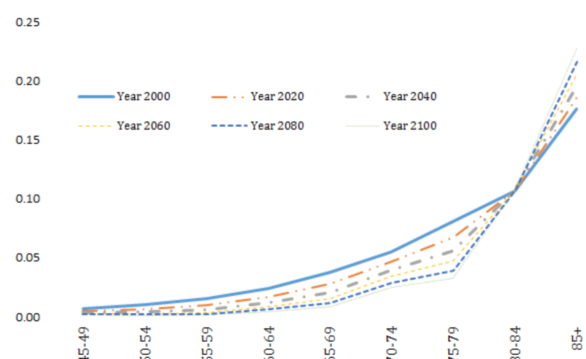
3.2 Mortality Forecast

In the previous subsection, we have reported the estimates of individual parameters of the forecast model. Now we report and reflect on fitted and forecasted mortality rates based on our adaptation of the LC model. The fitted mortality rates for 2000, and the forecasted ones for year 2020, 2040, 2060, 2080 and 2100 are shown in Figures 4 and 5, respectively in the linear and logarithmic scale. Note that since mortality at the younger ages are much lower, the scale in Figures 4a and 4b are chosen to be different to have greater visual clarity. The figures shows projected significant decrease in the infant mortality (age bracket 0–4); the decrease in mortality over time are gradual and for the middle ages, it is more visible in the logarithmic scale. For the higher age groups (55–79), mortality rates are projected to have significant improvement during the century.

Figure 4: Fitted and Forecasted Indian mortality

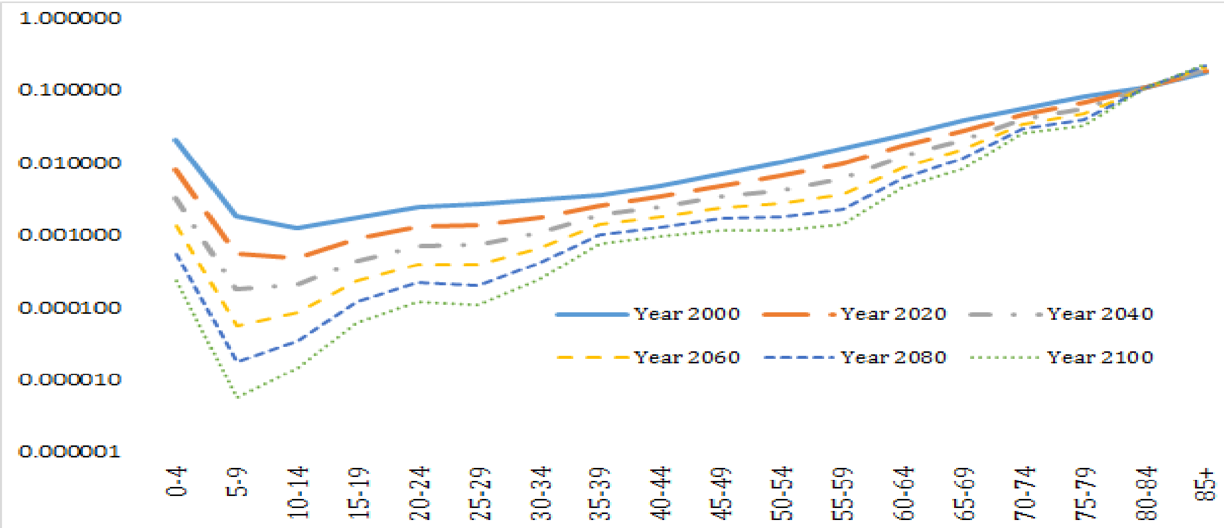


(a) Age brackets up to 44



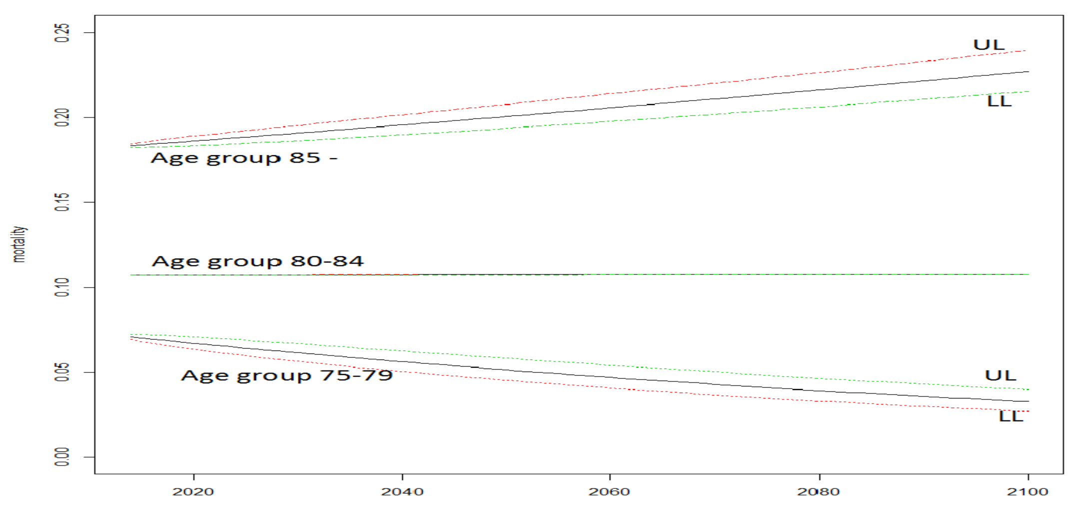
(b) Age brackets up to 45 onwards

Figure 5: Fitted and Forecasted Indian mortality in log scale



Naturally confidence intervals around the forecast are quite useful to assess reliability associated with the forecasts. To have greater visual clarity, not all projections are accompanied by associated confidence intervals in this article. To get some requisite idea, this is included in the Figure 6 below, which depicts the projected mortality (for total Indian population) in the three highest age-brackets, along with the upper limits (UL) and lower limits (LL) of the associated 95% confidence intervals. Focussing on the higher age-brackets is useful, since while as per model, the mortality rates reduce over time for most lower age-brackets up to 70–74, for the two high two highest age brackets, the mortality rates actually increase over time. This is intuitively justifiable and consistent with the discussion on observed sign of b_x values. For the age-group 80–84, the increase is very nominal (very small negative value for b_x, associated with very small standard error), and hence narrow confidence interval.

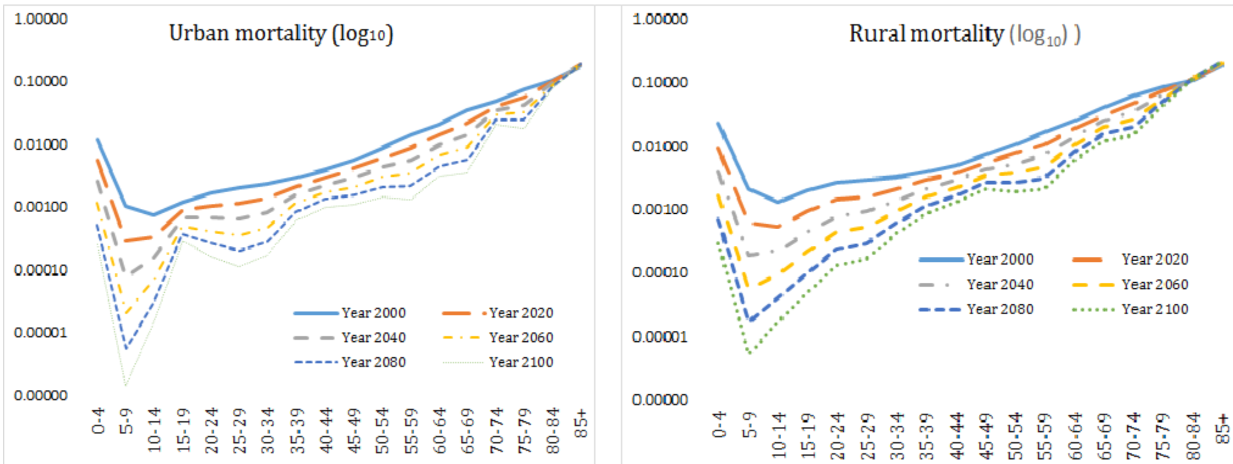
Figure 6: Forecasted mortality for the Total Indian population in the three highest age-brackets



3.3 A Comparison of forecasted Mortality: Urban-Rural and Male-Female

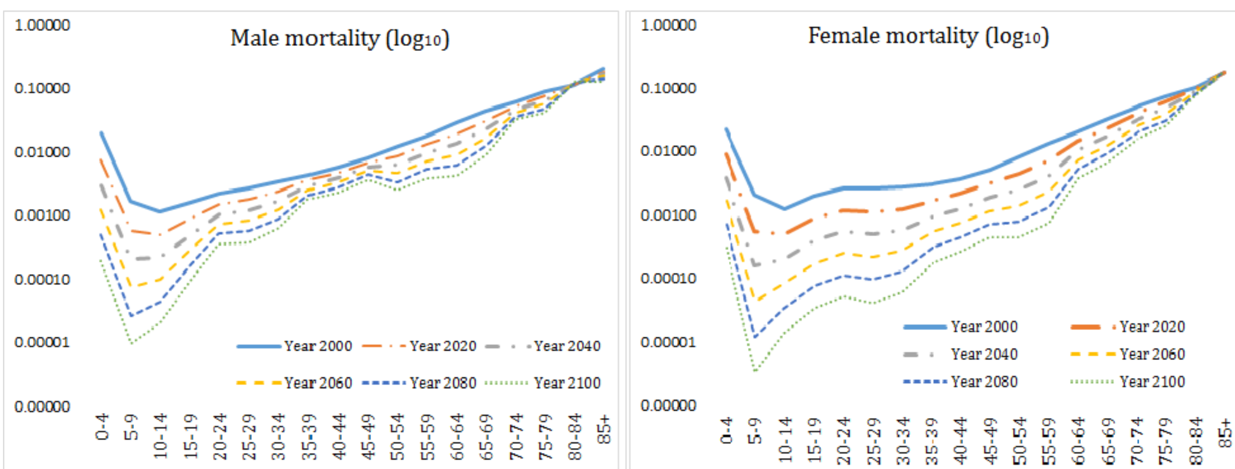
In Figure 7, we compare the projected mortalities, in gaps of 20 years, for the urban vis-a-vis rural population, in logarithmic scale. There is a great degree of commonality in the overall pattern and progression over time. One major difference is that the urban population (as compared to the rural) is projected to get increasing lower mortality in the age-group 5–9, but urban population is also projected to have higher mortality in the age-bracket 15–19, the possible cause of which has been speculated in the previous section, while discussing b_x.

Figure 7: Observed and forecasted Urban vs. Rural Indian mortality in log scale



In Figure 8, we perform the similar comparison between the male and female population. This predicts a substantially sharper fall in female mortality as compared to the male counterparts across most ages. These projections not only help us to build the population pyramids as demonstrated in the next section, but also projects a realistic possibility of achieving progressively greater gender imbalance in future.

Figure 8: Observed and forecasted Male vs. Female Indian mortality in log scale



Since the above comparative projected mortality rates are shown only at the interval of 20 years, in Figure 9, we exhibit the urban-rural comparison *across all years*, for three selected age-groups: 0–4, 25–29 and 55–59. For the last two age-groups, 0.025 is added to the mortality values for ease of visibility. The figure re-confirms that the urban mortalities are always lower, with the difference being substantial for the youngest age-group 0–4, with the gap narrowing with progression of time, for all the age-groups.

Figure 9: Forecasted Urban vs. Rural Indian mortality for select age groups

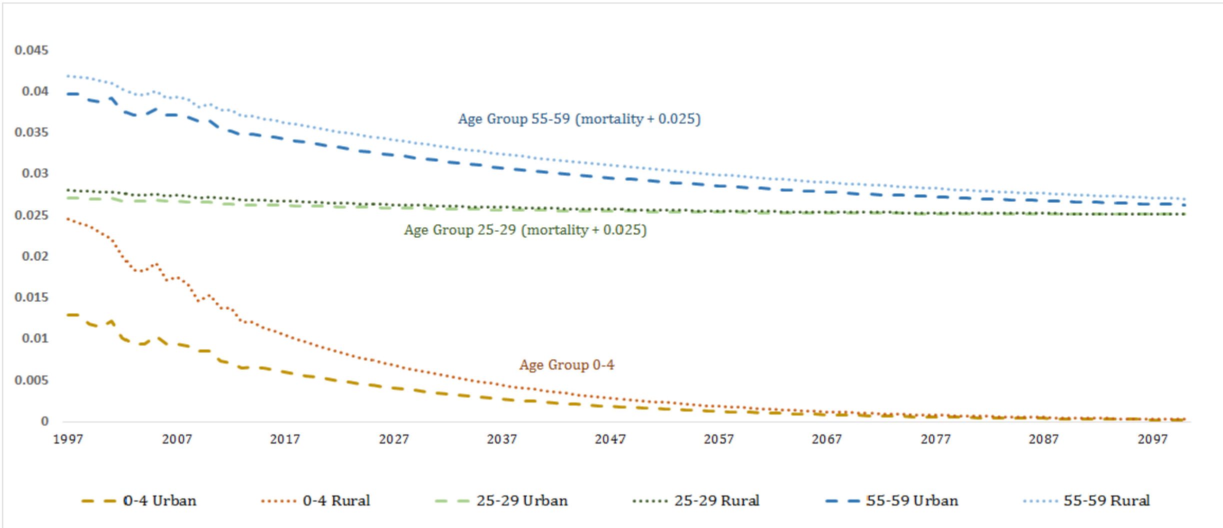
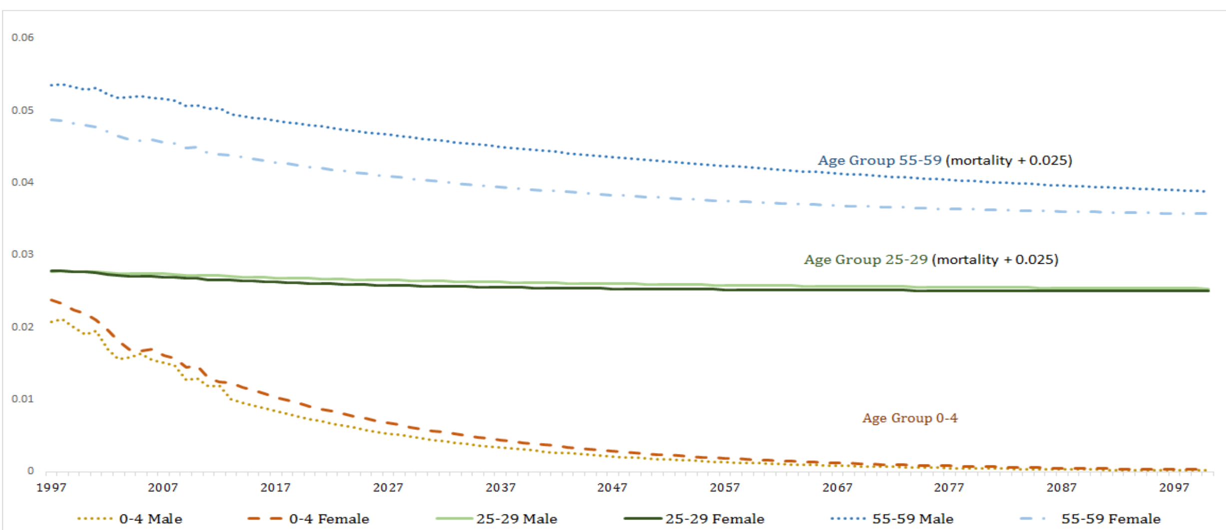


Figure 10 provides similar detailed comparison between male and female mortality. The biggest gender gap is for the age-group 55–59, with the female mortality being substantially lower. On the other hand, male mortality is marginally lower for 0–4 age-group, and the mortality rates are almost identical for the age group 25–29.

Figure 10: Forecasted Male vs. Female Indian mortality for select age groups

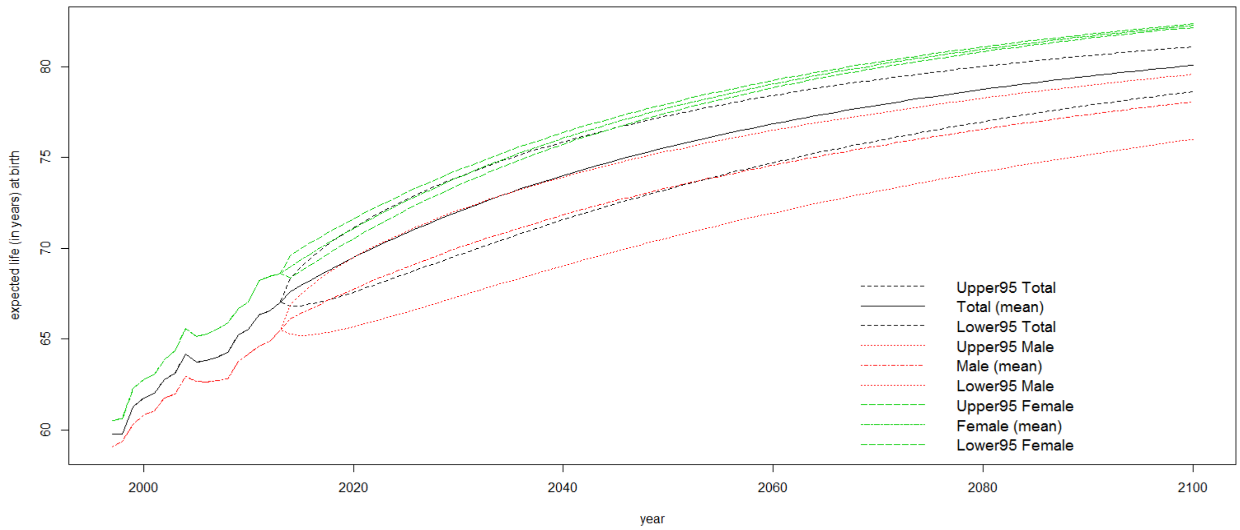


3.4 Projected Expected Life at Birth

One of the key utility of mortality calculations is in computing life expectation. While this can be done using routine actuarial calculation for the different ages, (i.e. expected remainder life, having survived a certain number of years), one of the most important among them is the expected life at birth. In Figure 11, we compare the expected life at birth for the total population and also compare between male vis-a-vis female subpopulation as a time series progression up to the end of the century. The projections are supplemented by corresponding 95% confidence intervals. The figure shows a steady expected increase in life expectancy from 60 to 80 over this century for the aggregate Indian population. The female life expectancy is more than male consistently, with the difference increase from about 1.5 years to 4.2 years during the century.

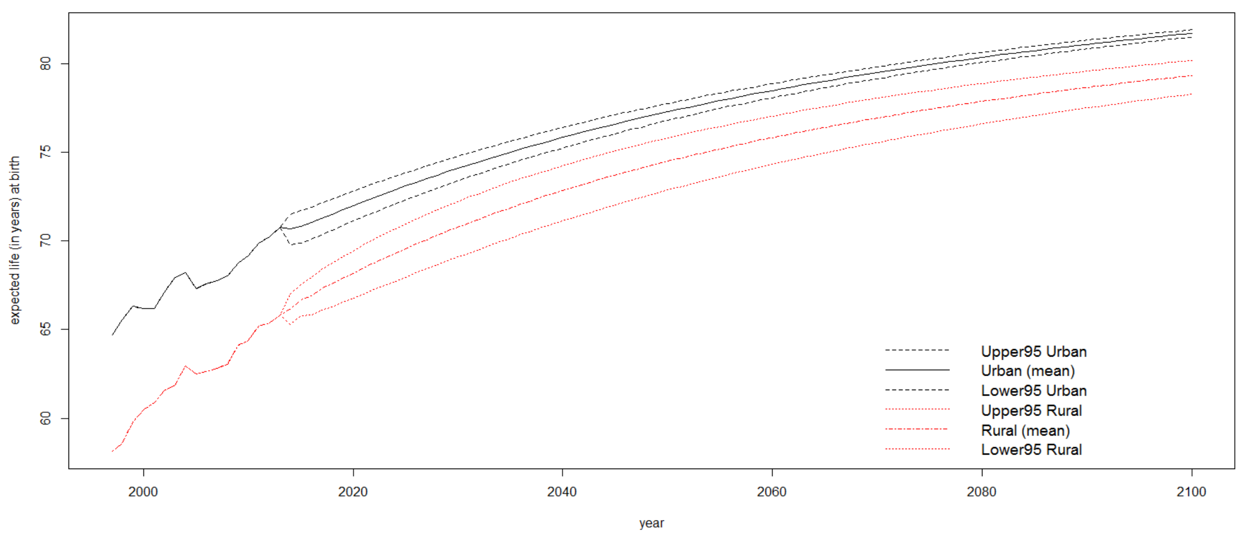
Interestingly, the confidence interval for the female sub-population is sharply and progressively narrower, as compared to the male.

Figure 11: Expected Life at Birth with 95% C.I.: Total, and Male vis-a-vis Female



The comparison of expected life at birth between urban and rural sub-populations is depicted via Figure 12. As expected, the life expectancy is uniformly higher for the urban compared to the rural sub-population; but the difference reduces from about 6.5 year to 2.4 years during the century. The confidence interval for urban population is also somewhat narrower than the rural, although the difference in that regard is not as substantial as in the female-male comparison.

Figure 12: Expected Life at Birth with 95% C.I.: Urban vis-a-vis Rural



It may be noted that the projected expected life numbers for the total Indian population, as well as the male and female sub-populations, broadly match with the corresponding projections in UN report [17].

4 Projected Population Demographics

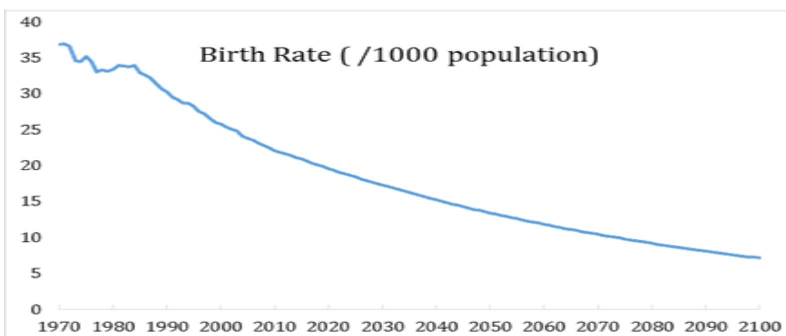
In this section we illustrate the application of the mortality modeling in projecting key population demographics. Other than projected mortality rates, three dimension of demographics play critical role, viz. (a) birth rate (b) gender ratio at birth and (c) migration rate. We first discuss these in Section 4.1. In Section 4.2, we project the population pyramids for the century. In Section 4.3, we focus on projected gender ratio. We also propose a new measure of gender imbalance that takes into account the difference across the age-groups. The projected picture of the measure is also highlighted. Finally in Section 4.4, we reflect on projected dependency ratio.

4.1 Related Demographics

Birth rate: Projection on birth rate, defined as number of children born in a year per thousand population, is a key to predicting future demographics like dependency ratio, population pyramid. An alternative, seemingly more attractive one, is to consider the fertility rate. However, in Indian context, the data for fertility rate is available for much shorter past period and more importantly, there are greater unreliability with this data, due to various data reporting/collecting issues, possible specific to India. Hence we choose to use the birth rate, in stead of the fertility rate.

Data on birth rate in India is published by the Ministry of Health and Family Welfare, and is available for the period from 970 till 2013; this data has been used adopting Box-Jenkins model with possible data transformation. Eventually, an ARIMA (0,1,0) model with drift was selected for the logarithm of the birth date, and on the basis of this, the birth rates are forecasted till 2100; this is shown in Figure 13.

Figure 13: Fitted and Forecasted Birth-rate in India



Gender ratio at Birth: The second critical component of constructing the population pyramid and other demographics is to forecast gender ratio at birth, defined as the number of girls born per 1000 boys born. Several studies have established significant inequality of gender ratio at birth across the globe. For example, a recent article in http://worldpopulationreview.com/countries/sex-ratio-at-birth-by-country/ reports this index in 2017 (although defined inverse to the definition adopted here) for two hundred countries and ranks them. Interestingly, not a single country has more girls born than boys; — this somewhat strange statistic is consistent with data published in many other earlier reports, even if not overtly publicized. Simple un-weighted average across the countries results in about 950 girls per 1000 boys. This statistic is particularly poor for the two most populous countries — China (870) and India (904) — which stand at the worst and fourth-worst position, respectively, on this front. This establishes that it would be unreasonable to assume gender neutrality at birth for any country, and certainly for India.

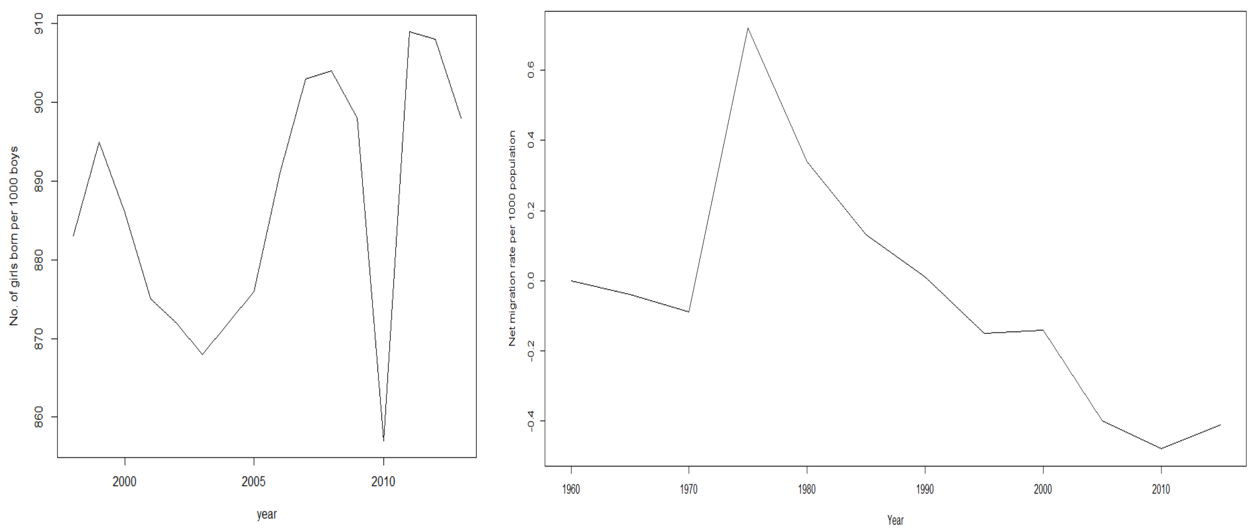
Report from Ministry of Home Affairs, Govt. of India, (ON668) and Ministry of Statistics and Programme Implementation Govt. of India (ON1168), provides the gender ratio birth figures for 1998–2013, this is shown in Figure 14a. No reasonable pattern seems to be emerging over the limited period. Formal exploration under ARIMA framework also confirms this by picking the white-noise model. Thus, while there are increasing awareness, campaign, governmental and social effort to address the possible human intervention in the observed gender ratio at birth, there is no observable improvement, at least as of now.

These considerations lead us to work under two possible projected scenario for the projected gender ratio at birth:

I The gender ratio at birth will remain constant at 900 (current level) in this century — this represents the likely scenario based on limited data, even if somewhat pessimistic.

II The gender ratio at birth will improve (increase) slowly at steady linear rate to reach 950 (global unweighted average) by the year 2100 — this represents somewhat optimistic scenario.

Figure 14: Gender Ratio at Birth and Net Immigration Rate in India



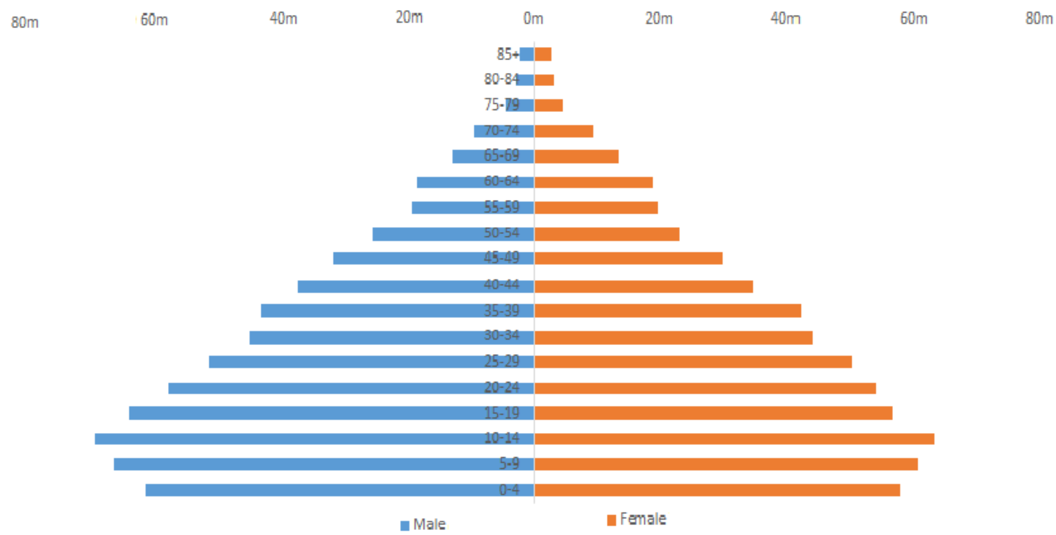
(a) Gender Ratio at Birth (b) Net Migration

Migration: Net migration, taking into account immigration as well emigration, can play an important role in the projection of population demographics for some countries. We have chosen to ignore this factor, primarily because the historical data shows the net migration has stayed at a very nominal level (-0.41 per 1000 in year 2015). (See, [11]). The time series on net migration is shown in Figure 14b. A secondary reason for ignoring this small factor is the lack of additional demographics in terms of age and gender for immigrant and emigrant population.

4.2 Projecting Indias Population Pyramids for the Century

A population pyramid is a diagram that provides a graphical illustration of the age distribution of a population, separately for the male and female, respectively in the left and and the right side of the diagram. It is essentially two back-to-back histograms (for male and female) of age-distribution, which nicely captures the age and gender distribution. India's population pyramid in 2011 is depicted in Figure 15.

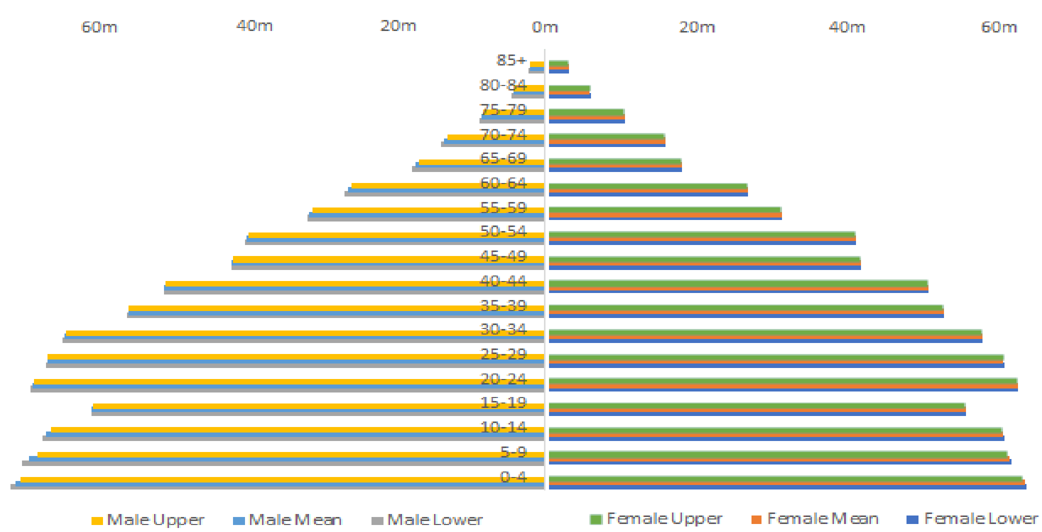
Figure 15: Population Pyramid in 2011



As a first step towards building the projected population pyramid, the LC model is extended onto mortality rates in India, as already shown in the previous section. Next, the forecasted birth rates are multiplied by the total population of the previous year to estimate the population added in that particular year. We have also built 95% confidence intervals around histograms of the pyramids based on lower and upper limits of the forecasted mortalities. In this manner, projected pyramids can be constructed for any of the years in the forecast horizon, for each of the two scenarios for future gender ratios at birth. For demonstration, we have chosen to show the projected pyramids for three of these years: 2025, 2050 and 2100.

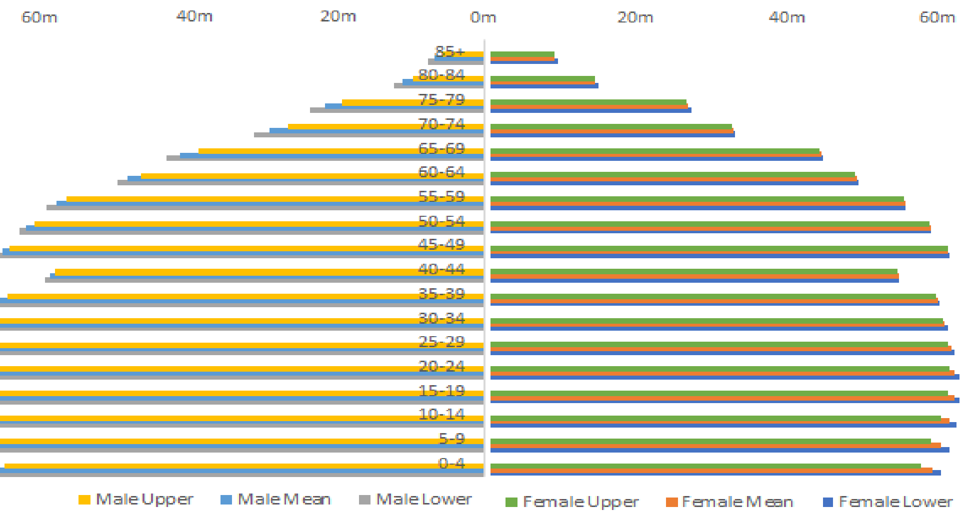
Figure 16 shows the predicted pyramids (along with 95% confidence intervals) for 2025, projected under Scenario I for gender ratio at birth.

Figure 16: Projected Population Pyramids in 2025: Scenario I



The predicted pyramids for 2050, projected under Scenario I, is shown in Figure 17.

Figure 17: Projected Population Pyramids in 2050: Scenario I



Next we look at Figure 18, which shows the projected population pyramid in 2100, under both the scenarios of projected gender ratio at birth.

Figure 18: Projected Population Pyramids in 2100

(a) Scenario I

(b) Scenario II

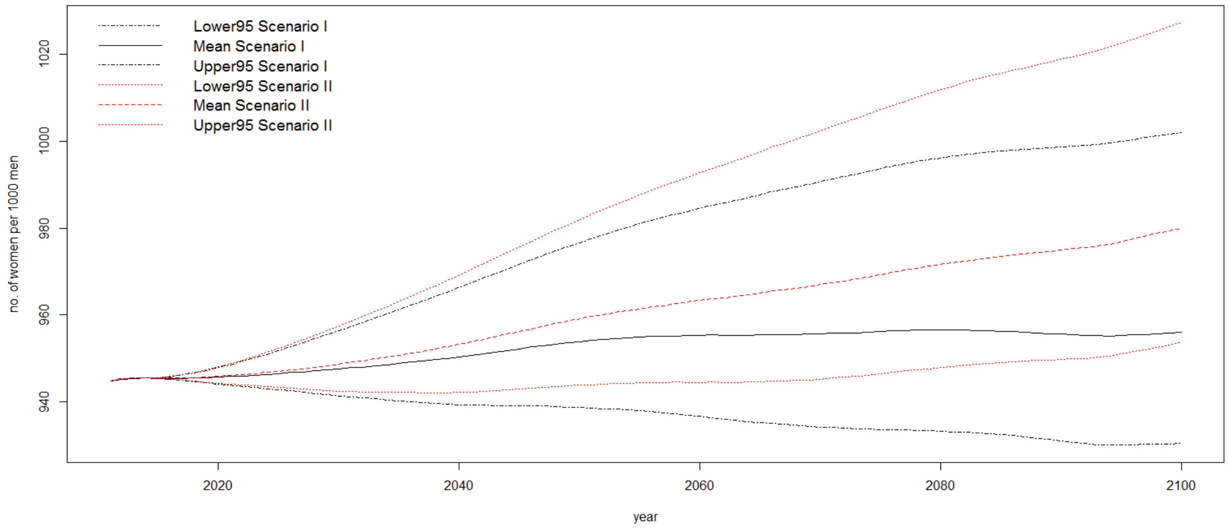
While there is some difference in the projected pyramids between the two scenarios, as we can see from Figure 18a and Figure 18b, it is quite nominal. Such differences are naturally even less for the year 2025 and 2050. Hence for brevity, we skip presenting pyramids for 2025 and 2050 under Scenario II.

The projected evolution of the population pyramids reveal several insightful information. There is gradual narrowing of base reflecting reduction in birth rate. The gradual bulge towards the top is reflective of reduced mortality across ages. The four pyramids show how India is likely to progress towards being like other aging population by the end of the century. The dip in the age-group 15–19, as remarked earlier, seem pronounced in the year 2025.

4.3 Gender Ratio and Measuring Age-specific Gender Imbalance

Using the model and projection of the previous sections we can construct and reflect upon the **gender ratio** (defined in India as the number of females per thousand male). The possible progression under the two scenarios (I and II, as defined in Section 4.1) is depicted via Figure 19. The 95% confidence intervals are built around the mean projected values.

Figure 19: Observed and Predicted Overall Gender Ratio with 95% C.I.: Under Scenario I and II



The above diagram shows the gender ratio, by combining all the age-groups. A more detailed plot, that shows the gender ratio age-group-wise, could be more insightful. This is given in in Figure 20. For visual clarity, only the mean values are depicted, without the confidence interval.

Figure 20: Observed Predicted Age-specific Gender ratio: Under Scenario I and II

(a) Scenario I

(b) Scenario II

While the overall gender ratio is a broad indicator of gender diversity or inequality, the difference in gender distribution may offset each other to a degree, across the different age-groups. For example, lower percentage of females at younger age-groups may offset higher percentage of females at the older age-groups, thus possibly showing greater overall balance between male and female, than possibly appropriate. Our projected gender-specific age distribution as depicted in the population pyramid, allows us to measure the age-specific gender imbalance. To address the overall gender-gap across the different age-groups, we propose the following gender-imbalance index:

$$GI = 2 \times \sum_x w_x \times \big(\max(f_x^M, f_x^W) - 0.5\big),$$

where w_x denotes the percentage of the population in the age-bracket x, while f_x^M and f_x^M respectively denote the percentage of male and female among people in the age-bracket x. The above gender-inequality index can be further subdivided into part which indicates the extent of male dominance and female dominance:

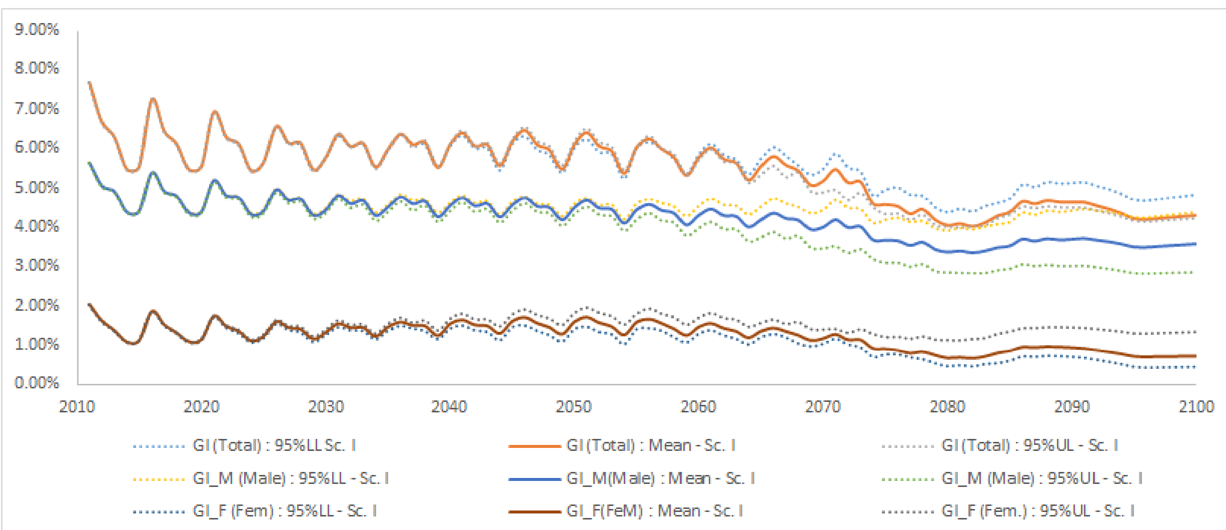
$$GI_M = 2 \times \sum_x w_x \times \big(f_x^M - 0.5\big) \times I_{f_x^M > 0.5}, \qquad GI_W = 2 \times \sum_x w_x \times \big(f_x^W - 0.5\big) \times I_{f_x^W > 0.5},$$

and naturally we have

$$GI = GI_M + GI_W.$$

In the above GI_M and GI_W denote the two components of the gender imbalance, respectively taking into account the parts where male and females are higher. But since only the *absolute* percentage differences are considered, they do not cancel each other.

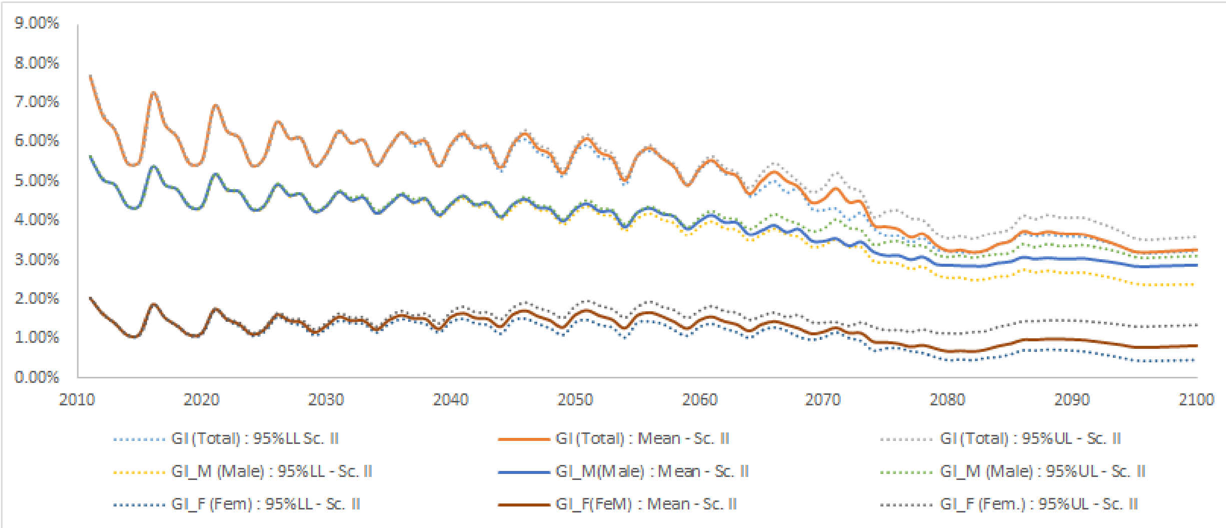
Figure 21: Observed and Projected Gender Imbalance: Under Scenario I



In Figure 21, we plot the projected values of this gender imbalance index under the scenario I (as introduced earlier in this section). Thus (even) under this scenario of gender ratio of birth staying at the present level throughout the century, the gender imbalance is projected to come down from 7.68% to significantly less than 5% by 2100 (95% confidence interval being (4.23%, 4.84%). Over the same period, the GI_M, the gender imbalance index, reflecting male dominance reduce from 5.64% to a mean level of 3.58% (confidence interval (2.88%,4.38%)), while the GI_F, the gender imbalance index, reflecting female dominance reduce from 2.04% to a mean level of 0.73% (confidence interval (0.46%,1.35%)). Thus, we project a greater gender balance by the turn of the century, even if the gender ratio *at birth* does not improve during this period. Naturally, this may be explained by higher rate of decrease in female mortality as compared to the male subpopulation.

The same under scenario II is shown in Figure 22. Because of assumed improvement in the gender ratio at birth under this scenario, the reduction in gender imbalance is more striking with the overall gender imbalance projected to reduce to a level in between (3.22%, 3.60%) by 2100, while the gender imbalance reflecting male dominance projected to come down to (2.39%, 3.12%), and the imbalance reflecting female dominance projected to come down to between (0.48%, 1.35%).

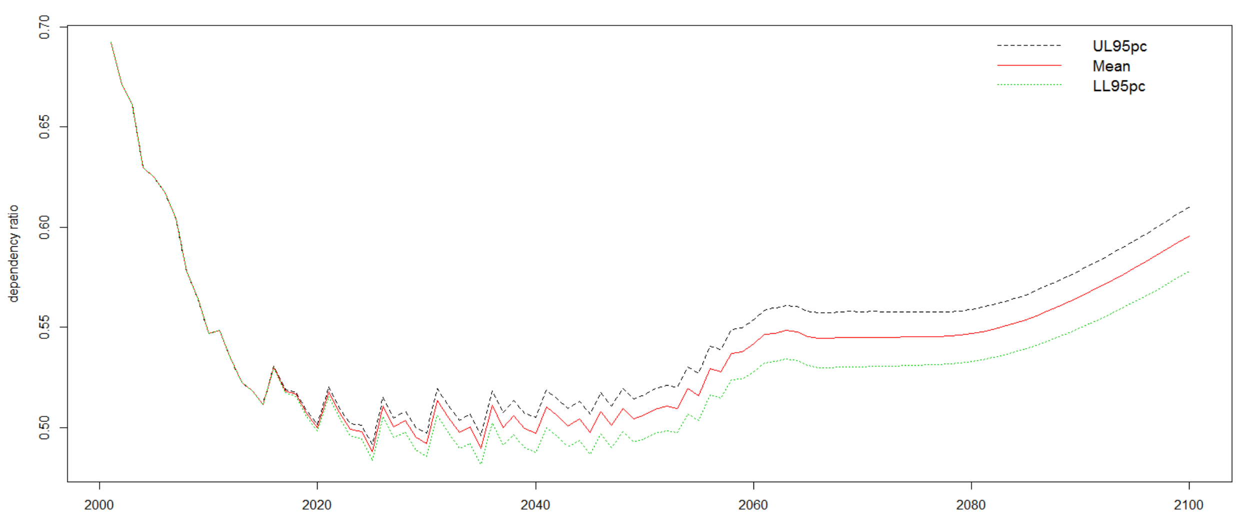
Figure 22: Observed and Projected Gender Imbalance: Under Scenario II



Hence we predict that that the gender imbalance is going to reduce during this century, in any case, although the extent may be dictated by actual gender ratio at birth. Male dominance is the significant component of the gender-imbalance and typically women dominance component small and stagnant over time. Also the gender imbalance reflecting women dominance are projected to be more or less identical under the two scenarios of gender ratio at birth, while that is not so for the part reflecting male dominance. The cyclical pattern of periodicity 5 years in the index is due to the mortality rates being considered in block of five years. Under both the scenario, the large part of the imbalance is contributed by few age-groups only.

4.4 Projecting Dependency Ratio

Figure 23: Observed and Projected 95% Confidence Interval of the Dependency ratio



The dependency ratio is defined as the ratio of (combined) population in age-groups up to 14 years and exceeding 65 years to that in the age-group 15–64. The 15–64 age-group is considered to be the working population, and hence the dependency ratio provides information on average number of people dependent on a person who are in the working age. A high dependency ratio may be considered to be a potential liability of an aging population. It is of interest to forecast and reflect upon the dependency ratio based on our mortality model; this is done in Figure 23, along with the 95% confidence interval. From the figure, we note that the dependency ratio is not projected to go up in spite of projected reduced mortality. As a cautionary remark, we note that the critical explanation for the pattern in dependency ratio is the assumed the birth rate reduction model (which offsets reduced mortality rates).

As the plot of the projected dependency ratio shows, the ratio of the dependent population to the working age population is on a decreasing trend from 70% in 2000 to a significantly lower level at around 50% till 2020 and it is likely to stay around the 50% mark till 2050. The dependency ratio is projected to go up marginally to around 55% by 2060 and the again stay essentially at that level till 2080. This may free up resources and boost economic growth for a fairly long period without having too much adverse effect in terms of financial dependency due to increase in mortality. By 2100, when the dependency ratio would be increasing to a level of around 60%, this increase would be largely contributed by the increase in the elderly population. This may enable a transition to a mature consumption driven economy.

The dependency ratio, as projected here, matched quite closely with [17] till 1960; however the cited reference projects a sharper increase in dependency ratio in the longer term (after 1960). The difference is possibly due to the difference in methodology used.

5 Summary and Concluding Comments

We have implemented the Lee-Carter model to a country with sparse data – India. While the the data paucity suggests caution should be exercised, especially while using the long term forecasts based on the adopted model, the predicted mortality rates of multiple sub-classes in the country provide many useful directions and insights, over all forecast horizons. In the work, we have focussed on projected life expectancy, population pyramid, gender diversity and dependency ratio.

Developing a measure of age-specific gender imbalance and its projection is a contribution of the present work. An significant contribution of the work is the LC model based projection of the population pyramid. Once mortality rates had been predicted, we also predicted the birth rates of the total population; then utilized the current population to project the population pyramid of the country in the future.

Comparison with published mortality rates: The Insurance Regulatory and Development Authority (IRDA) publishes the Indian Assured Lives Mortality (IALM) rates in the form of actuarial tables. These published mortality rates matches reasonably with the LC fit, although the former rates are lower, with the difference being substantial for the two lowest age-brackets as well as in the higher ages. Such differences are to be expected since, since mortality rates reported in actuarial tables are based on insured population which is only a subsection of the entire population. The study of the past differences, along with the projected mortality rates can actually be used by the policymakers to decide on the timing of up-dation of the actuarial table. This is important especially for India, as these actuarial tables have been are updated very infrequently and after varying time gaps (three past updates were made in 1970-73, 1994-96, and 2006-08).

SVD vs ML in LC: In the present work, we primarily implemented the LC model using SVD as in the pioneering work and many follow-up adaptations. However it is also popular to use the ML method instead of SVD. We present a brief summary of fit from the two methods. In implementing ML based LC model, we followed the methodology of Brouhns et al. (2002). Because of the data format requirement, we had to use a slightly different source/format of data — downloaded from World Health Organization (WHO) web-site, [19]. Here the Indian data was available from 2000–2015, and only separately men and women.

Table 3: LC model fit/accuracy – SVD vis-a-vis ML

	SVD		ML	
Population	% variability explained	MAPE	% variability explained	MAPE
Indian Male	99.93%	3.61%	99.99%	6.42%
Indian Female	99.98%	5.37%	99.99%	7.02%

Limitation: The LC model is adapted in Indian context on the basis of very limited horizon of 17 years of data. Naturally the forecast for longer period, as considered in the work, is appealing from the perspective of utility, the projections are subject to a fair deal of uncertainty especially when they are done for the latter part of the century.

Future work: Naturally it is of interest to apply the model recursively as progressively more data becomes available. Currently we have used constant mortality within particular age-block period. Naturally, this can be modified using linear interpolation or other smoothing techniques. This was not so appealing given limited data, and is planned to be explored in future.

Methodologically, we wish to explore imposing non-negativity constraint on b_x parameters in SVD implementation of the LC model, along the lines of Das & Sen (1994). This is already done in the literature on ML adaptation of LC model – however, in the present context, it was not enforced, as we believed that the mortality are increasing for the age-brackets. It remains to be seen if this is a manifestation of short-term feature only, as speculated by Lee & Carter (1992). At any rate, this may be explored by recursive adaptations of implementation of LC model with data from other countries, starting from similar early stage of data availability.

Concluding Remark: The population pyramid shows that India currently enjoys a significant demographic dividend; with much of the population in the lower age groups of 0-30. However, some experts anticipate gradual prominence of a second demographic dividend in India before the onset of the century. When the country may face a situation of falling working age population and a growth in the elderly (60+) population - which is referred to as the second demographic dividend. As additional resources would be necessary for taking care of an elderly-concentrated population, emphasis must be given to facilitate savings in the long run; and shift the economy to be more service-oriented that can sustain the second dividend. Rate of GDP growth of an economy is positively correlated with the rate of growth of the labor force. India is poised to take benefits of two population dividends, subject to successful implementation of carefully promulgated policies making use of both the dividends. Labor demand has to be nurtured by facilitating skill development and motivating higher investment in the secondary and tertiary sectors. Emphasis should be given to encourage the working class to facilitate savings and investments that would cause a sustainable boost to the economic development. The present study may contribute towards objective assessment of the above features and recommendations.

References

[1] Berkum, F.V., Antonio, K., & Vellekoop, M. (2013). Structural changes in mortality rates with an application to Dutch and Belgian data. Leuven : Kath. Univ. Leuven, Fac. of Business and Economics. AFI 1379.

[2] Booth, H., & Tickle, L. (2008). Mortality Modelling and Forecasting: a Review of Methods. *Annals of Actuarial Science.* 3: (1-2): 3–43

[3] Brouhns, N., Denuit, M., & Vermunt, J.K. (2002). A poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics,* 31: 373–393.

[4] Cairns, A.J.G., Blake,D., Dowd, K., Coughlan, G.D., Epstein, D., Ong,A., & Balevich,I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal,* 13(1):1–35.

[5] Lee, D.R., & Carter, L.R. (1992). Modeling and Forecasting U. S. Mortality. *Journal of the American Statistical Association*, 87 (419): 659-671.

[6] DÀmato V., Piscopo G., Russolillo M. (2011), The mortality of the Italian population: Smoothing techniques on the Lee-Carter model. *The Annals of Applied Statistics* 5(2A): 705–724.

[7] Das, S. & Sen, P.K. (1994). Restricted canonical correlations. *Linear Algebra and its Applications*, 210: 29–47.

[8] Manisha Dubey, I. I. (2014). Trends and Prospects of Mortality by Age and Sex in India: 1991-2030. PAA 2014. http://paa2014.princeton.edu/papers/141995

[9] Gaille S (2012), Forecasting mortality: when academia meets practice. *European Actuarial Journal* 2: 49–76.

[10] Ghosh, S.K., and Sriram, K. (2005) Gender and Smoker Distinct Mortality Table Development. https://www.actuariesindia.org/downloads/gcadata/7thGCA/GSDMT_S%20Ghosh%20&%20K%20 Sriram.PDF

[11] India - Population - Net migration rate. https://knoema.com/atlas/India/topics/Demographics/ Population/Net-migration-rate?action=export&gadget=indicator-preview-host

[12] Mason, A.a. (2006). What Is the Demographic Dividend? Finance and Development - IMF.

[13] Makridakis, S., Wheelwright, S.C. & Hyndman, R.J. (2005) Forecasting: Methods and Applications. John Wiley & Sons (Asia) 3rd edition

[14] Indian Labour Statistics (2012–13) . Shimla / Chandigarh: Government of India - Ministry of Labour and Employment.

[15] Sample Registration System, Census India, Ministry of Home Affairs, GOI, http://www.censusindia.gov.in/2011-Common/Sample_Registration_System.html

[16] Sample Registration System Baseline Survey (2014). http://www.censusindia.gov.in/vital_statistics/BASELINE%20TABLES07062016.pdf

[17] United Nations, Department of Economic and Social Affairs, Population Division (2015). World population prospects. The 2015 revision. https://esa.un.org/unpd/wpp/Graphs/DemographicProfiles/

[18] Wang, J.Z. (2007). Fitting and Forecasting Mortality for Sweden: Applying the Lee-Carter Model. Mathematical Statistics, Stockholm University Examensarbete 2007:1, http://www.math.su.se/matstat

[19] Global Health Observatory data repository. World Health Organization. http://apps.who.int/gho/data/view.main.60740?lang=en

[20] Yasungnoen, N., & Sattayatham, P.(2016). Forecasting Thai Mortality by Using the Lee-Carter Model. Asia-Pacific J. of Risk & Insurance 10(1): 91-105.