"Systematic Bias[...] in U.S. Social Security Administration Forecasts," or is There?

Jack Luby

5/3/2019

Abstract

Through their 2015 article, "Systematic Bias and Nontransparency in U.S. Social Security Administration Forecasts," (Kashin, King, and Soneji (2015d)) authors Konstantin Kashin, Gary King, and Samir Soneji present evidence of a persistent downward bias in U.S. Social Security Administration life expectancy and expenditure forecasts. In a companion paper, "Explaining Systematic Bias and Nontransparency in US Social Security Administration Forecasts," (Kashin, King, and Soneji (2015a)) they expand upon this argument, presenting evidence that private and public pressures, in concert with entrenched nontransparency, have bred consistent bias toward low-cost projections. As a key part of their argument, Kashin et al. argue against SSA forecasting methodology, which makes use of hundreds of subjectively selected UROD (Ultimate Rates of Decline) for various demographic groups and causes of death. Through the selection of these parameters, they argue that the intrusion of bias is inevitable and contend the necessity of creating models which are less influenced by the subjective opinions of SSA actuaries. In analysis of their argument, I propose five rudimentary models and compare their performance against historical SSA forecasts. I find that SSA projection "biases" are largely attributable to the similarities between their forecasts and a simple OLS time series regression. While the performance of other models vary widely across samples, SSA and linear models perform consistently. I propose an alternative hypothesis to explain the apparent "bias" in SSA forecasts. I argue that "biases" seen in the works of Kashin et al. are the result of 1) a commitment by the SSA to produce consistent forecasts in line with regressions toward a linear trend of life-extending progress, 2) an unexpected and unpredictable boom in life expectancies from 2000-2010, and 3) the authors use of "true" life expectancy data which is distinct from that used by the SSA and diverges from it after 2000.

Background

The Social Security Administration (SSA) is a key administrative agency of the United States federal government which is tasked with administering Social Security, a social welfare program consisting of retirement, survivors, and disabled social insurance. Funded through the impostion of taxes on more than 166 million working Americans, these benefits are paid from the Social Security Trust Fund. This fund provides a key form of insurance to those U.S. residents who may be unable to work. The balance of the Social Security Trust Fund is therefore of paramount importance in ensuring the long-term health and safety of the U.S. non-working population.

Beyond the administration of these benefits, the SSA is tasked with understanding and projecting the future health of the trust fund as it relates to key variables such as fertility, mortality, and tax rates. These variables establish the size of U.S.'s working population, retiree population, and the supply of funds which are available to be paid from the former to the latter. As such, understanding the present and future states of these variables is key in ensuring the long-term health of the trust fund and insurance of millions of non-working Americans. The SSA publishes their projections for these values, and more, in the annual OASDI Trustees Report.

In recent years, much attention has been paid to the faltering health of the Social Security Trust Fund. Deservedly so. According to the SSA's own 2017 forecasts, the trust fund balance is expected to fall to 0 by

the year 2029. Through their two companion 2015 papers, authors Konstantin Kashin, Gary King, and Samir Soneji assert an even bleaker reality. They argue that the trust fund is actually in greater danger than is suggested by these SSA projections and that, because policy decisions relating to payroll and Social Security taxes are directly guided by the Trustees Report, preservative policies may yet come too late or be insufficient to counteract the true current unhealth of the program.

The Problem

Through their companion 2015 papers, Kashin et al. argue that implicit biases in Social Security forecasting methologies have precipitated inaccurate projections of key demographic and economic variables. These biases, which have crept in through the subjective determination of hundreds of Ultimate Rates of Decline (UROD) of mortality, have caused the underestimation of life expectancy forecasts, and therefore an overestimation of the trust fund's long-term health.

The below figure visualizes this perceived problem. Each annual Trustees Report analyzed in this paper contains within it a projection of expected period life expectancy in the year 2010, a value representing years left to live for the average person given current mortality rates. In this figure, we see that deviations from truth for each published Trustees Report, especially following the year 2000, are significant. Hovering around a full year of error, these projections appear significantly skewed. Even the 2010 Trustees Report, which only sought to project life expectancy one year into the future, fails to represent true values within a half a year for any of the four groups. This problem is yet worse in consideration of the fact that alongside these 'Intermediate Cost' projections, 'Low Cost' and 'High Cost' alternatives were proposed to represent forecasting uncertainty. These values are often represented by the SSA as siblings to the upper and lower bounds of a 95% confidence interval. However, we find that for the year 2000 and on, not a single true value falls below the 'High Cost' SSA estimate. As they describe in their work, this fact could have major implications on efficient policymaking.

It is with these issues in mind that Kashin et al. assert the necessity for sweeping changes in Social Security forecasting methodologies. Their first recommendation calls for 1) full transparency of data and forecasting procedures and 2) the simplification of models so that they easily understood and resistant to bias. This paper takes on these recommendations through the implementation of several rudimentary modeling techniques. A description of those models is below. The performance of these models sheds an interesting light on the "bias," or lack thereof, in Social Security Administration models.

Forecast Error of Life Expectancy in 2010 by Year of Report
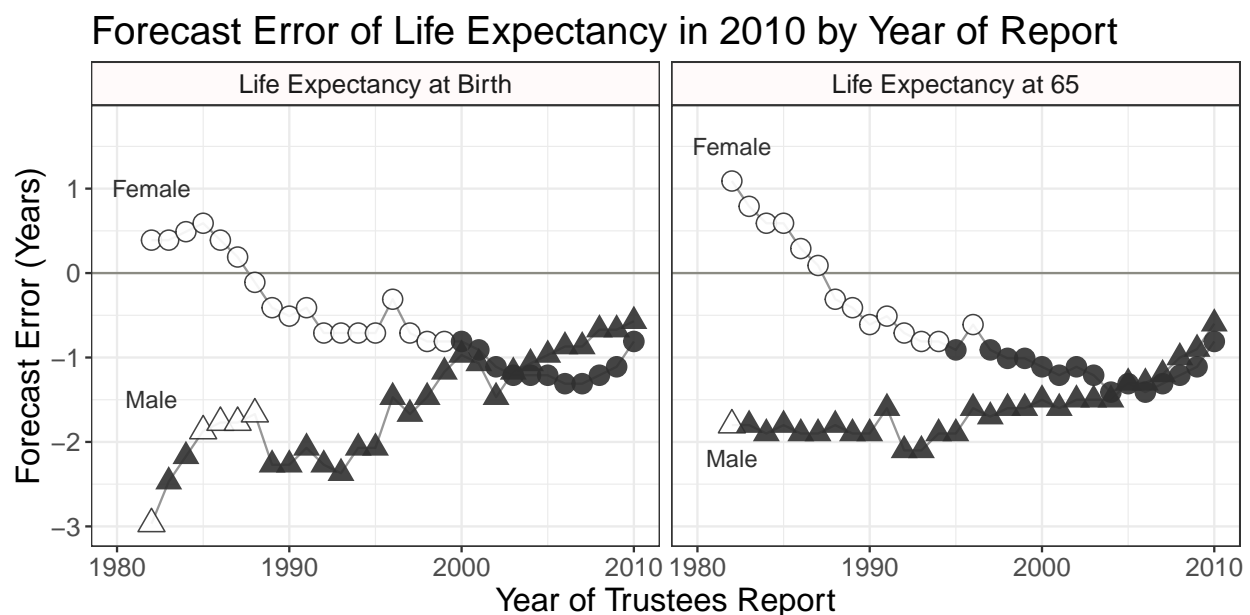
Data

The data used in this paper were obtained from four sources. First, all data and code necessary to replicate the authors' original papers were obtained from those papers' respective Harvard Dataverse repos (Kashin, King, and Soneji (2015c)) (Kashin, King, and Soneji (2015b)). Second, Social Security projections from the year 2011 were obtained from the 2011 Trustees Report (("Table V.a3.- Period Life Expectancy" 2011)). Third, true life expectancy data for the years 2011-2017 were obtained from the Human Mortality Database (HMD) ((n.d.)). Finally, health data regarding obesity, smoking, and addiction rates for use as external regressors were obtained from online National Health Interview Survey (NHIS) data (Lynn A. Blewett and Williams (2018)). All data used in this paper are digitally published and available to the public.

The models used in this paper are fit to life expectancy data acquired from the Human Mortality Database. These models will be assessed upon their performance relative to Social Security Administration forecasts. For fair comparison, all models are fit only to data which would have been available to the Social Security Administration at the time of their forecasting.

Preregistration

As noted by authors Brian A. Noseka, Charles R. Ebersole, Alexander C. DeHavena, and David T. Mellorain in their recent paper "The Preregistration Revolution," the vital differentiation between prediction and postdiction is often recognized in theory but ignored in practice (Nosek et al. (2018)). While research purports to first propose hypotheses then test their veracity, this order can be easily inverted. As a result, overfitting and hindsight biases can easily make their way into research, obscuring the truth. In order to prevent the intrusion of these biases, preregistration of model parameters has become common. Under preregistration, all models are explicitly declared prior to any data analysis. In this way, retroactive model alterations are prevented and bias is limited. As this paper centers around the removal of bias from life expectancy forecasting, all models proposed herein were digitally preregistered in the Government 1006 Slack prior to accessing any future SSA, HMD, or NHIS data.

Models

Five models are fit to the data. The first four are univariate, meaning that they only make use of trends in the life expectancy data itself to make forecasts. The fifth model integrates external regressors. The first model is a simple univariate linear time series forecast fit using base `stats` package in r. The next three make use of the forecasting packages `forecast`, `prophet`, and `ets`, respectively, to analyze the same univariate time series. The final model also uses the `forecast` package, integrating several external regressors to account for specific health trends.

Linear Model

Making use of the r function `lm()`, a linear model is fit to Human Mortality Database life expectancy data for the years 1982-2010. Taking a traditional form `Y = a + bX`, a line is fit which minimizes the sum of the model's squared errors (that is, the square of the difference of the observed and fit values). The trend of this model is used to project future values for the years 2011-2017 such that we expect life expectancy in the year 2011 to be `b` greater than the fit 2010 value, and so on. This model implies a case in which there is some constant amount of life-extending progress made annually around which true values fluctuate.

Non-linear Univariate Models

Three non-linear univariate models are fit to the same HMD life expectancy data for the years 1982-2010. Despite increased modeling complexity, these models still operate only by extrapolating trends in the raw life expectancy data. These models primarily differ from the linear model by giving greater weight to more recent observations. They may also attempt to capture seasonality trends should non-linear patterns appear in the data.

auto.arima()

The first non-linear model makes use of the r function `auto.arima()`, automatically fitting an appropriate ARIMA model to the data. ARIMA stands for AutoRegressive Integrated Moving Average. The 'Autoregression' component refers to the model's use of lagged values of the same variable's past values (in this case, past life expectancies). The 'Integrated' component refers to the differencing of values from their past values, making values stationary and allowing for more accurate forecasting. Finally the use of a 'Moving Average' smooths short-term fluctuations and captures trends in the data by taking the mean values of subsets of the data and analyzing trends in those averages. Three additional parameters, one for each of these components, are used to capture seasonal trends. By searching over all possible models within preset constraints, this function returns the best ARIMA model according to the Akaike information criterion (AIC), with an additional correction for use in data with small sample sizes (such as ours) (Dalinina (n.d.)).

prophet()

The second non-linear model makes use of the r function `prophet()`. Prophet is an open source forecasting procedure created by Facebook's Core Data Science Team. This procedure forecasts based upon the fitting of non-linear trends with daily, weekly, and yearly seasonality. As stated in their own documentation, and as would be expected given the forecasting needs of a company like Facebook, these forecasts perform best when given large quantities of data with strong seasonal effects (("Prophet Is a Forecasting Procedure Implemented in R and Python. It Is Fast and Provides Completely Automated Forecasts That Can Be Tuned by Hand by Data Scientists and Analysts." n.d.)).

ets()

The third non-linear model makes use of the r function `ets()`, automatically fitting an appropriate exponential smoothing model to the data. `ets()` models are fit similarly to `auto.arima()` models. Their primary difference, however, lies in the distinction between moving average techniques and exponential smoothing. In moving average models, like those fit in `auto.arima()`, all observations are weighted equally in forecasting. Exponential smoothing, conversely, makes use of exponentially decreasing weights for older observations. Recent trends will therefore tend to influence forecasts more than they would an ARIMA model (("Forecasting: Principles and Practice," n.d.)).

External Regressor Model

The final model again makes use of the `auto.arima()` function from the `forecast` package. In this case, however, external regressors are added to the model so that known health trends can be accounted for in the model more accurately. The two external regressors added are the obesity rate (percentage of population with BMI > 30) and the smoking rate (percentage of population who have smoked more than 100 cigarettes in their life and have smoked in the past 30 days) for each respective age and gender group. So that these forecasts are not made on data which would not have been available to the SSA, data beyond 2010 are forecasted using the ets() function.

Performance

Model projections are plotted below against true values. This visualization gives a general understanding of the differences between models.

The relative performances of the models are judged primarily on the mean squared error (MSE) of their forecasts. This value is calculated by squaring the difference of each years' forecast and true value. Models with a lower MSE are therefore more accurate to the data.

Over a 5-year timeframe, all models outperform SSA forecasts in at least two of the tested demographic groups. All but one model (prophet) return an average MSE lower than that of the SSA forecasts. Two forecasts, the linear model and the ARIMA model with external regressors, outperform SSA forecasts in all tested demographic groups.

Over a 7-year timeframe (up to the most recent data), the general MSE performance trends hold. However, in the 'Male Life Expectancy at Birth' demography, SSA forecasts perform better than all other models. This trend can be seen visually in the figures below, as all fit models tend to vastly overestimate life expectancy in this group. SSA forecasts also perform relatively better over this longer timeframe, and interestingly appear similar in trend to the linear model, although their lower intrercept negatively impacts their MSE. Further analyses of model performances are presented below.
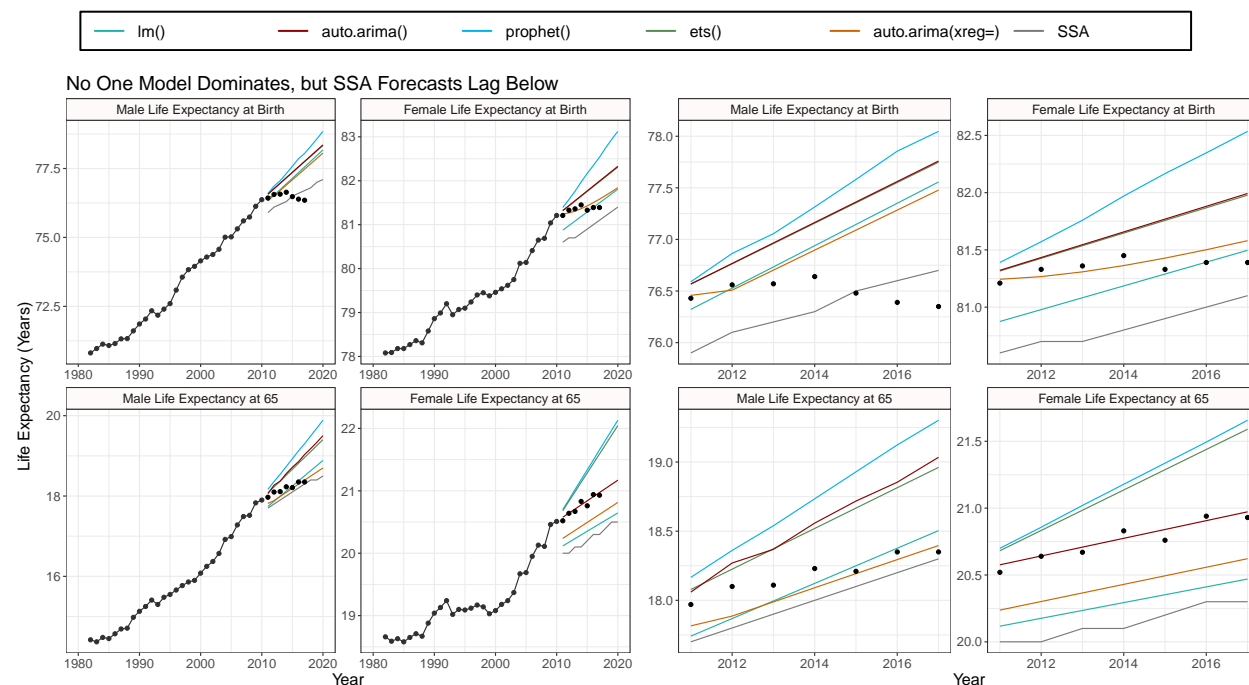
| type	ssa	linear	prophet	ets	arima	arima_xreg
Male Life Expectancy at Birth	0.14908	0.1143361	0.4044395	0.2492616	0.2546024	0.0918780
Female Life Expectancy at Birth	0.36240	0.0769123	0.2440779	0.0546142	0.0586294	0.0050029
Male Life Expectancy at 65	0.05440	0.0264549	0.2124742	0.0778463	0.0939221	0.0207773
Female Life Expectancy at 65	0.37028	0.2042751	0.1311332	0.1067880	0.0028341	0.1036515

type	ssa	linear	prophet	ets	arima	arima_xreg
Male Life Expectancy at Birth	0.1302857	0.4216715	1.0072665	0.6502723	0.6619082	0.3617483
Female Life Expectancy at Birth	0.2926000	0.0565551	0.4926990	0.1210195	0.1281362	0.0105265
Male Life Expectancy at 65	0.0424286	0.0223926	0.3667984	0.1398005	0.1703098	0.0155814
Female Life Expectancy at 65	0.3797000	0.2161671	0.2131082	0.1741424	0.0024434	0.1084639

Linear Model

A simple linear regression performs well over both the 5-year and 7-year timeframes, even outperforming SSA forecasts in all four demographies over the 2010-2015 timeframe. Especially when compared to the failures of the non-linear models, the success of this model suggests that there may be some constant trend at which life expectancy rises annually for each demographic group. While true values jitter around this linear projection, life-extending technology and health trends may cause a linear upward life expectancy trend. According to this model, male life expectancy at birth will tend to increase by approximately .206 years per year, female life expectancy at birth by approximately .104 years per year, male life expectancy at age 65 by approximately .127 years per year, and female life expectancy at age 65 by approximately .059 years per year.

The strong performance of this model is in line with past research on life expectancy trends. Much research has indicated an upward linear trend in life expectancy, although some research has suggested this trend was disrupted around the year 2000 (Lee (2019)). This model may indicate a regression back to this linear trend.

Performance of Simple Linear Model
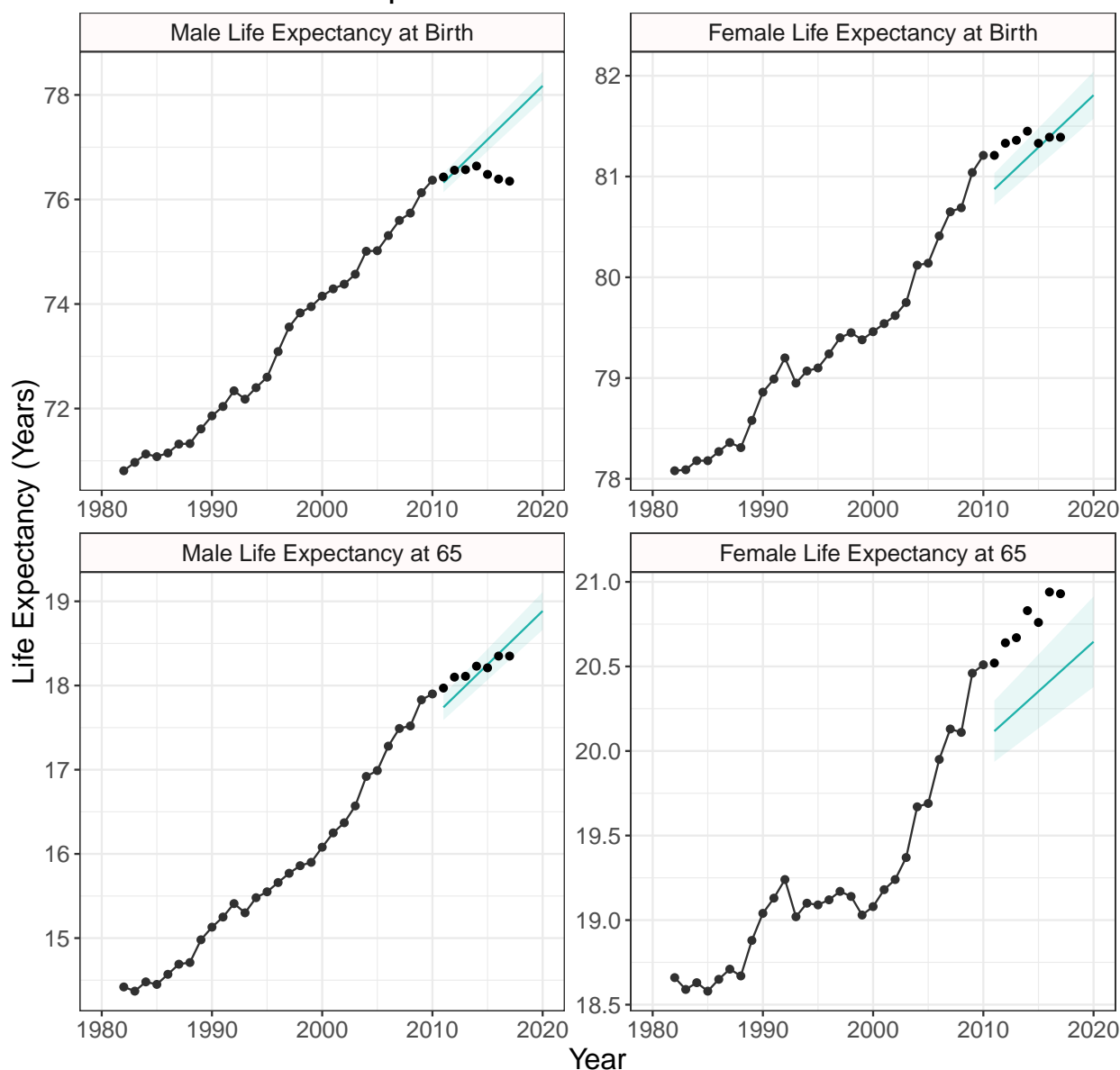
Nonlinear Models

The nonlinear univariate models tested perform relatively poorly, especially in the case of the 'Male Life Expectancy at Birth' subgroup. These models, as described in the 'Models' section above, attempt to fit nonlinear trends to the data. These fits are therefore more dependent upon recent trends than those of linear regressions. This can be seen in the figure below. Nonlinear univariate models appear to overfit to recent trends. In this case, our models overfit to large increases in life expectancy for all subgroups from the years 2005-2010, extrapolating these high trends to future years. As life expectancies stall somewhat in all subgroups after 2010, our projections overestimate true values, resulting in poor fits.

Observed Period Life Expectancy and Nonlinear Univariate Model Performances
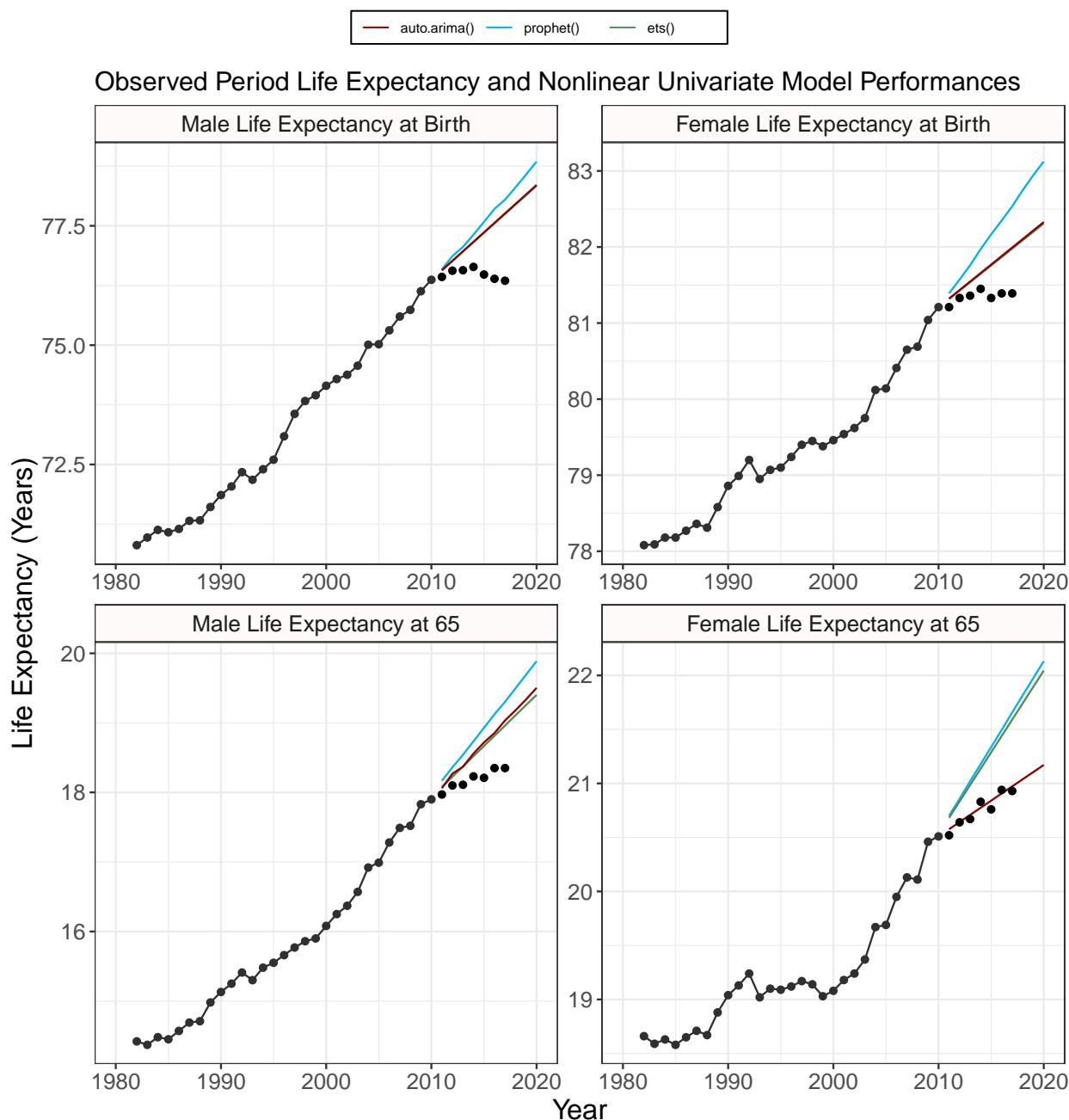
External Regressor Model

The ARIMA model with external regressors performs relatively well, as demonstrated by its low MSE across subgroups. This model appears to capture the general trends of the data, including the dampening of upward trends after 2010. It appears that by integrating external health data on smoking rates and obesity rates, this model has better predictive power than our univariate models. While obesity rates have continued to increase somewhat linearly over time, this model may be capturing slowing of the downward trend in smoking rates in the years leading up to 2010 (See: Appendix). Making use of this information, the model accurately projects a dampening of the observed upward trend in life expectancies. Further, this model's 95% confidence intervals capture 19/20 true values in the 2011-2015 timeframes. This is better than any other model and indicates that this model could provide meaningful guidance in short-term policymaking.

Performance of Arima Model with External Regressors
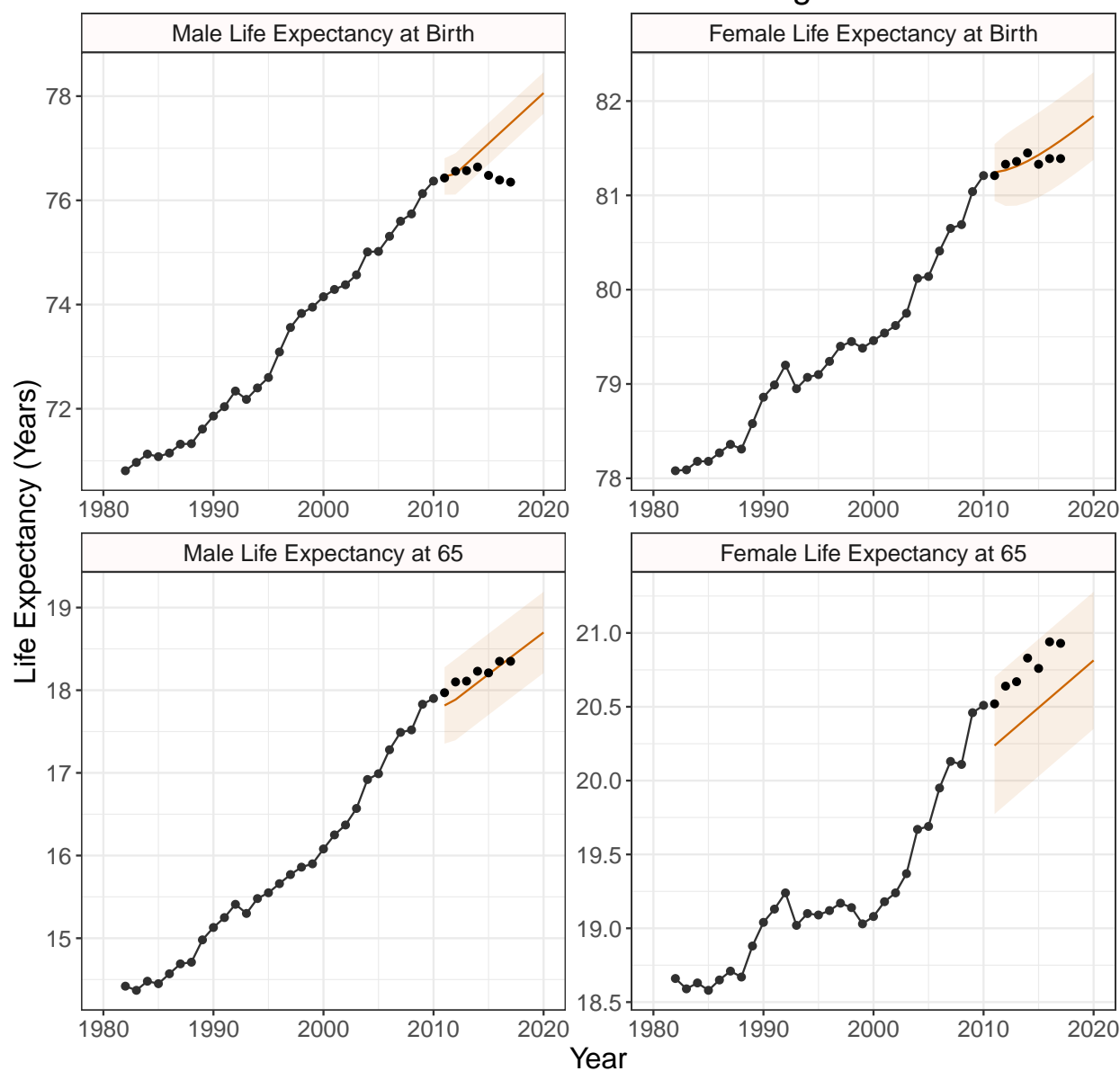
SSA Model

The Social Security Administration's forecasts perform poorly, but in several cases perform better than would be expected given the biases displayed in the work of Kashin et al. While these forecasts certainly tend to underestimate true values, the forecasts for later years much more closely approximate them and, importantly, contain the true values within their 'High Cost' and 'Low Cost' estimates in three of four 2015 cases. This would not be expected given the biases discussed in the companion papers upon which this work is based.

Of note, the trends of these SSA forecasts closely approximate those of the linear model fit above. This data suggests that the inflexiblity of the Social Security models which Kashin et al. discuss is not an inherent weakness in their forecasts, but rather a strength. Much as the linear model suggests some constant level of progress to which life expectancies will regress, so might these Social Security Administration forecasts. For this reason, despite their inability to project the boom in true life expectancy values in the 00's, these forecasts are somewhat successfully predictive of the regressions seen in the years after 2010. The ability of linear models to accurately project life expectancy over time, supported by past research, is captured by the consistency of the SSA forecasts (Lee (2019)).


Performance of SSA Forecasts
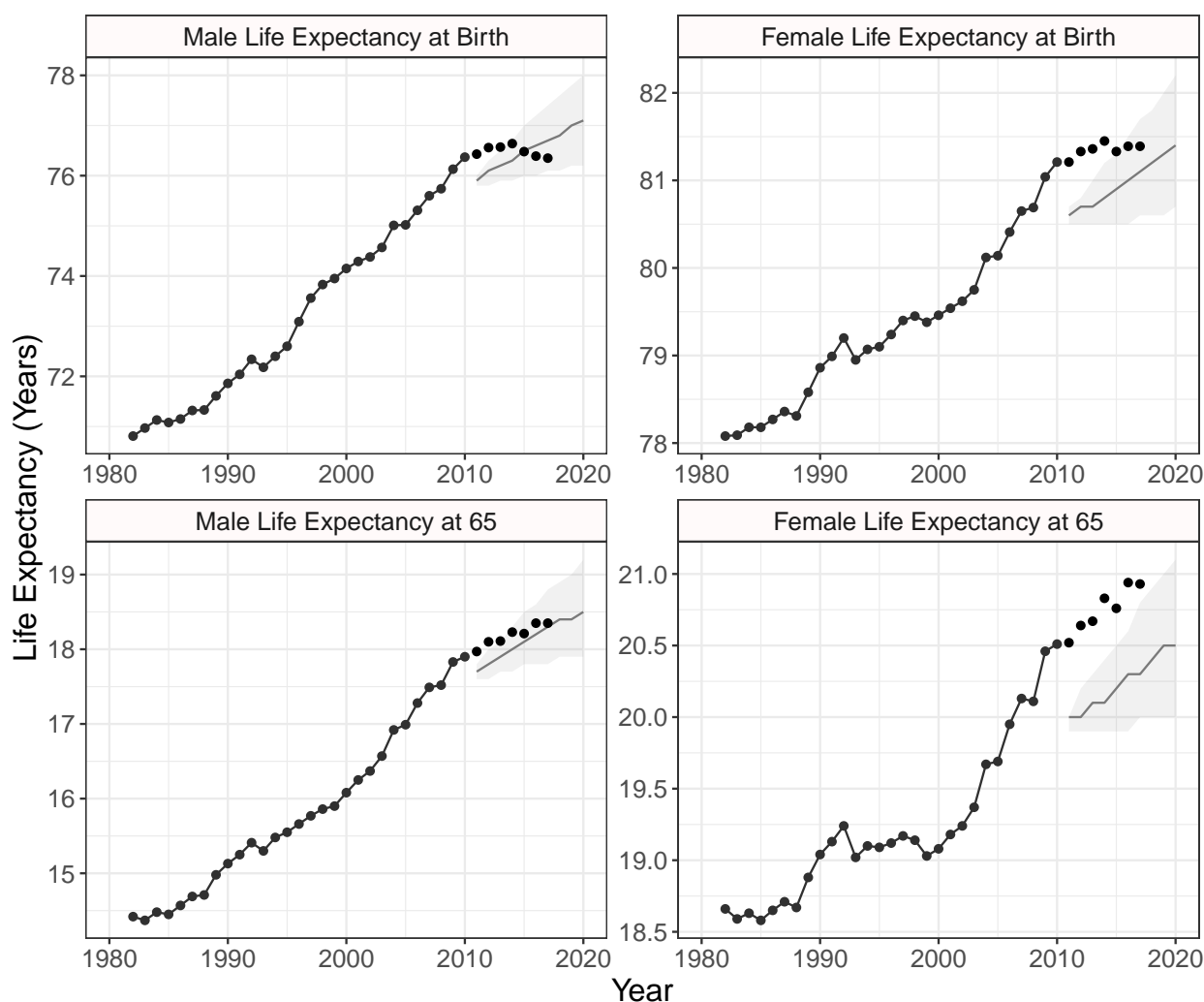
Further Analysis

As a further analysis of the linear and ARIMA with external regressor forecastss relative to the SSA's, models are refit to pre-2000 and pre-2005 data to forecast subsequent 5-year life expectancies. The fit of these models is compared to the 2001 and 2006 Trustees Reports, which respectively offered forecasts, with high and low cost estimates, for the years 2005 and 2010. Once again, SSA forecasts tend to underestimate true life expectancy values. In 6 cases of 8, SSA forecasts perform worse than ARIMA with external regressors. And in all 8 cases, they perform worse than a simple linear regression.

The fact that the SSA forecasts outperform the ARIMA model in half of the 2001-2005 subgroups is, however, revelatory of an important point: life expectancies boomed unexpectedly in the 00's. Neither of our models, which performed best when analyzing 2011-2015 data, were able to capture the increased upward trend of life expectancies during this time period. We could criticize the SSA's actuaries for their inaccuracies, as done by Kashin et al., or we could recognize that projecting forward leaps such as these might be nigh impossible.

Our best hope of projecting future values might require an expectation of regression to some constant level of progress, as suggested in our linear models, and as appears to be suggested by these SSA forecasts. Looking at the 2011-2017 data above, this expectation of regression may not be misguided.
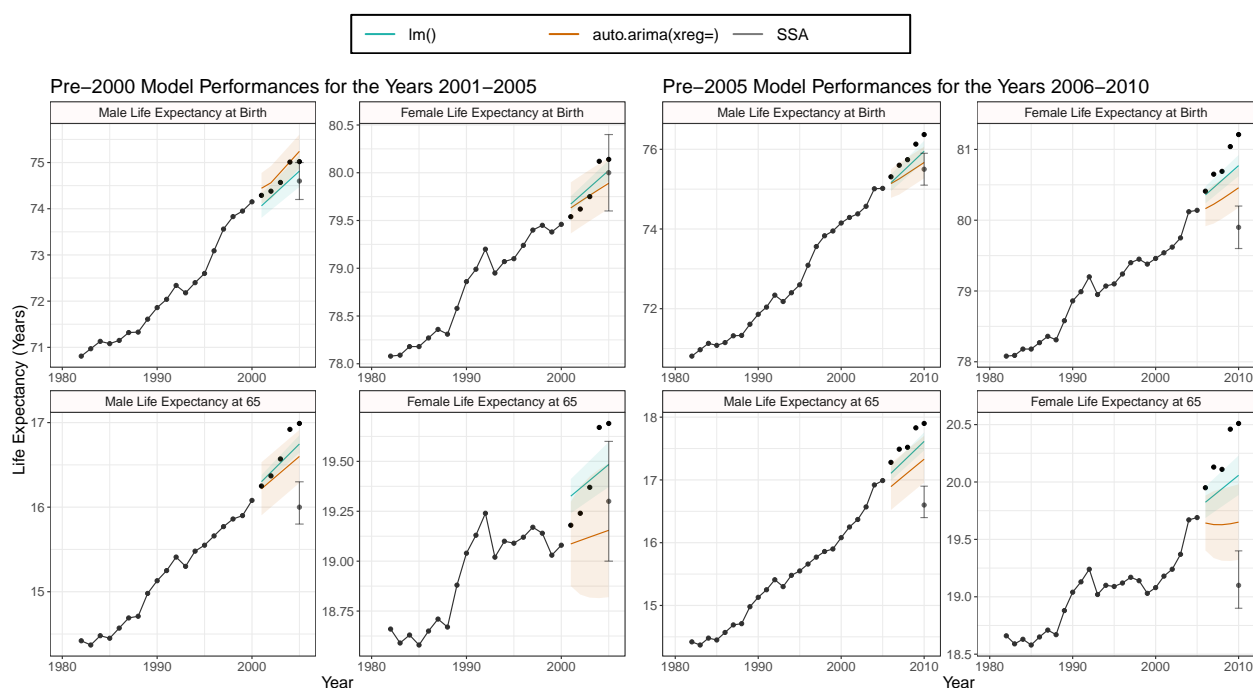

Data Considerations

While it seems as though linear models tend to outperform SSA forecasts, we must also note the stunning similarities between the two. As shown in the figure below, the trends of these projections are near identical. They are differentiated by their intercepts. It is vital that we consider the data against which we compare the SSA forecasts.

While Kashin et al. wave their hand at the differences between Human Mortality Database and Social Security Administration life expectancy forecasts, stating that "The small differences in estimated life expectancy between the two sources do not account for the much larger error rates and patterns reported in this article," the figure below demonstrates significant differences between the two opening up, interestingly, from the years 2000-2010.

The Social Security Administration founds its forecasts on its own recorded life expectancy values, meaning that we would expect their forecasts to fall below HMD values by at least the differences displayed below. Given that these differences have opened up in recent years, their forecasts are likely to be even further biased downward from HMD data. Were they to found their forecasts on HMD data, they would likely look quite similar to the linear models above, which have already been established to perform well in forecasting future values.

Data Discrepencies Between HMD and SSA Observed Life Expectancies
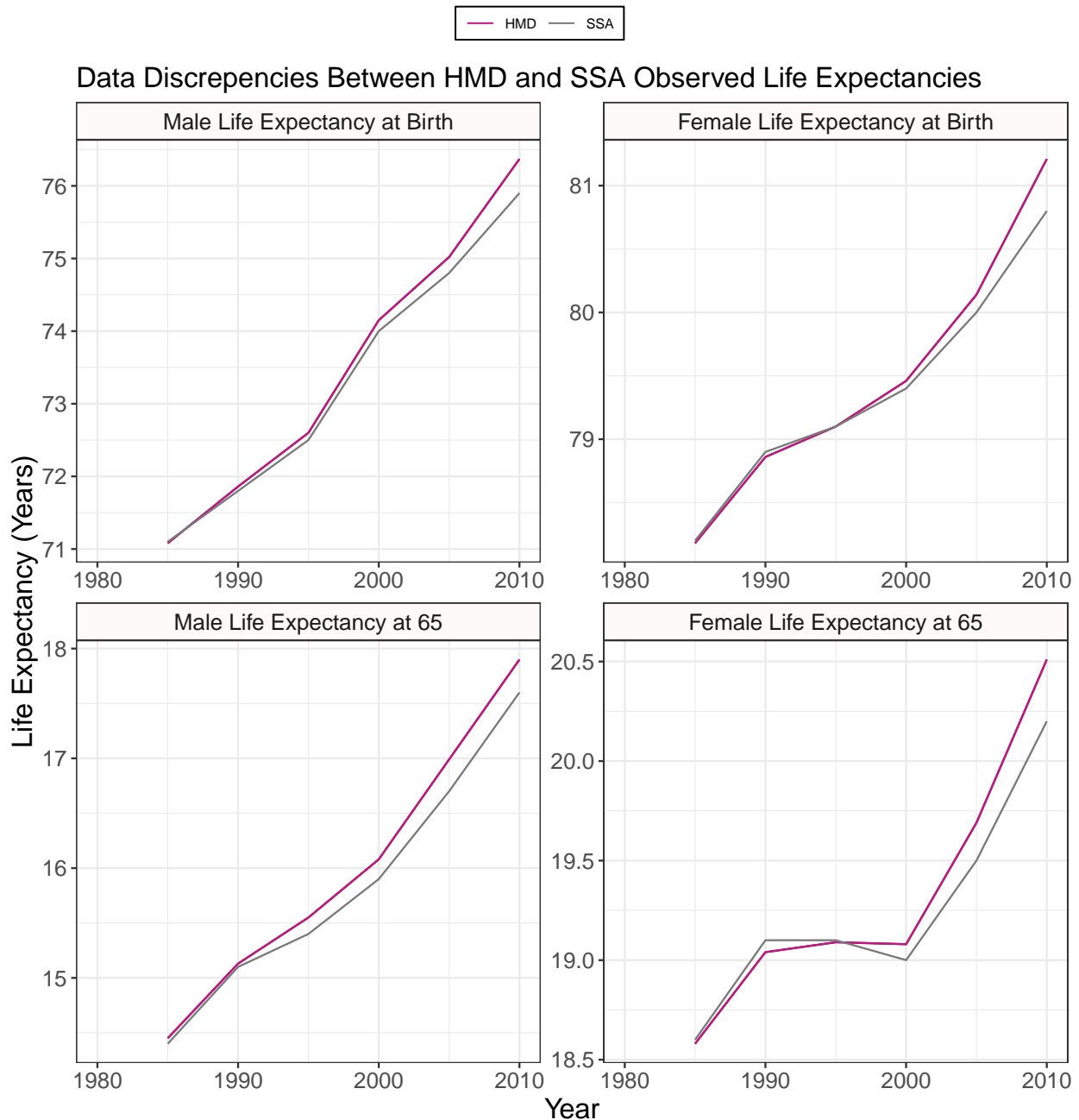
Discussion

I find that the biases highlighted by Kashin et al. in their companion 2015 papers are not nearly as significant as asserted. This finding is based upon three observations: 1) The SSA's models closely mirror simple

OLS regressions. The SSA is committed to the creation of consistent forecasting. To this end, they have produced near linear forecasts which appear to be intentionally uninfluenced by temporary trends. Past 2010, true life expectancy values regress toward linear projections. This suggests that SSA forecasting methods may be well-guided. 2) Kashin et al.'s paper analyze a time period in which an unpredictable boom in life expectancies occurred, causing all models tested to appear biased downward. 3) Kashin et al. primarily make use of Human Mortality Database observed life expectancy values, while the Social Security Administration bases their forecasts on their own data. After the year 2000, these two values diverge with SSA recorded values dipping well below HMD values. These three factors have likely created much of the perceived "bias" suggested by the authors, rather than any faults or biases in forecasting methodology.

References

Dalinina, Ruslana. n.d. "Introduction to Forecasting with Arima in R." *Oracle DataScience.com.* https://www.datascience.com/blog/introduction-to-forecasting-with-arima-in-r-learn-data-science-tutorials.

"Forecasting: Principles and Practice." n.d. *8.10 ARIMA Vs ETS.* https://otexts.com/fpp2/arima-ets.html.

Kashin, Konstantin, Gary King, and Samir Soneji. 2015a. "Explaining Systematic Bias and Nontransparency in Us Social Security Administration Forecasts." *Political Analysis* 23 (3): 336–62. http://pan.oxfordjournals.org/lookup/doi/10.1093/pan/mpv011.

———. 2015b. "Replication data for: Explaining Systematic Bias and Nontransparency in US Social Security Administration Forecasts." Harvard Dataverse. https://doi.org/10.7910/DVN/28323.

———. 2015c. "Replication data for: Systematic Bias and Nontransparency in US Social Security Administration Forecasts." Harvard Dataverse. https://doi.org/10.7910/DVN/28122.

———. 2015d. "Systematic Bias and Nontransparency in Us Social Security Administration Forecasts." *Journal of Economic Perspectives* 29 (2): 239–58. https://www.aeaweb.org/articles.php?doi=10.1257/jep.29.2.239.

Lee, Ronald. 2019. "Mortality Forecasts and Linear Life Expectancy Trends." In *Old and New Perspectives on Mortality Forecasting*, edited by Tommy Bengtsson and Nico Keilman, 167–83. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-05075-7_14.

Lynn A. Blewett, Risa Griffin, Julia A. Rivera Drew, and Kari C.W. Williams. 2018. "IPUMS Health Surveys: National Health Interview Survey." IPUMS. https://doi.org/10.18128/D070.V6.3.

Nosek, Brian A., Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor. 2018. "The Preregistration Revolution." *Proceedings of the National Academy of Sciences* 115 (11). National Academy of Sciences: 2600–2606. https://doi.org/10.1073/pnas.1708274114.

"Prophet Is a Forecasting Procedure Implemented in R and Python. It Is Fast and Provides Completely Automated Forecasts That Can Be Tuned by Hand by Data Scientists and Analysts." n.d. *Prophet.* https://facebook.github.io/prophet/.

"Table V.a3.- Period Life Expectancy." 2011. Social Security Online. https://www.ssa.gov/OACT/TR/2011/lr5a3.html.

n.d. Human Mortality Database, University of California, Berkeley (USA); Max Planck Institute for Demographic Research (Germany). www.mortality.org or www.humanmortality.de.