12 Comparative Analyses

We now turn from illustrations of our approach to direct comparisons of forecasting performance with other methods. We know from theory alone, and from many simulations in practice, that our method will substantially reduce forecast mean square error, relative to existing methods, when all four of the following conditions hold:

1. We observe multiple time series of observations, each of which is relatively short, sparsely observed, or noisy.
2. Different observations on our outcome variable are related in informative ways, such as through the known similarity of expected mortality in neighboring age groups, successive time points, nearby geographic areas, or interactions among these.
3. We observe informative covariates.
4. Patterns in future data are similar to or predictable from those in our observed data.

Under these conditions, our forecasts will easily outperform those based on separate least-squares regressions applied for each time series (due to nonlinearities or large variances); approaches without covariates, such as linear extrapolation from a random walk with drift or Lee-Carter methods (due to omitting informative covariates); standard time-series, cross-sectional econometric approaches (due to invalid parameter constancy assumptions and the requirement that covariates be observed for all cross sections or be omitted); and classical Bayesian approaches that smooth based on coefficients (because prior information is available about expected mortality, not coefficients).

Our approach will not necessarily outperform existing methods if all four conditions do not hold. In mortality forecasting in particular, conditions 1 and 2 hold generally. All forecasting methods assume condition 4 and, even if it is false, forecasters continue if only to provide the best systematic summary of our knowledge about the future based on knowledge we have available now. Thus, the key to valid forecasts, in the area where we are able to do something, is condition 3, informative covariates. Until now, many of the covariates that have been collected for mortality forecasting have been relatively weak predictors, primarily because previous forecasting methods worked only for covariates available for *all* cross sections. Because this constraint is not necessary with the methods we propose, new data collection strategies can proceed, and separate variables can be collected in each country, age group, cause of death, and sex that are most appropriate for each area. If the new technology we make available is used, then these new data collection efforts have the potential to improve the accuracy of mortality forecasts.

Making it possible to marshal the power of so many more informative and context-specific covariates is a key advantage of our approach. Because forecasting better by having better covariates is such an obvious result, we do not pause to demonstrate its real value, even though it will often be the biggest reason our method produces better forecasts in practice. Instead, we make comparisons in this chapter that illustrate the advantages of our priors, keeping the (relatively uninformative) covariates the same.

Yet, the simplicity of how to make the comparison is deceiving. In fact, what this method should be compared to is not clear at all. So-called standard methods, such as the Lee-Carter method, were not designed to deal with the type of problems that this method is supposed to help to solve, namely sparse and noisy time series. Also, usually the performance of any method can be improved by performing additional preliminary analysis, carefully estimating the parameters or the covariates to be used, or adding extra steps (such as an intercept correction). How much effort should spent improving each method to obtain a fair comparison is a question raised by Hand (2006) and Efron (2001) in the context of comparing classification algorithms.

Here, we perform some relatively simple experiments that help to answer reasonable questions many researchers would ask. For example, even if in most instances it is not fair to compare this method to Lee-Carter's, most people would still like to see some sort of comparison anyway, at least as a reality check. Also, it is reasonable to ask whether this method performs better than least squares, and how much better. In the next section we present the results of a few comparisons. We perform the comparisons on countries in our database with more than 25 observations for the all-cause mortality time series in the period 1950–2000. We use all observations up to year 1990 as our sample and test the forecasts on the period 1991–2000. All the covariates have been lagged 10 years, so when we write GDP_t we really mean GDP_{t-10}. Because the out-of-sample period is quite short, we cannot expect, on average, huge differences among the methods, given that all forecasts are quite smooth in such a short time period. We perform analysis of different types, rather than performing a very detailed analysis of one case, to emphasize that there are many ways of performing comparisons and different ways of using our methods.

We begin with data with strong patterns, observed over long periods of time and with few missing data points. Because such an approach violates condition 1, we do not expect our methods to add much value. We then move to the much more common mortality data, where condition 1 holds and our approach has more to add.

12.1 All Causes in Males

We start with all-cause mortality, which is one case in which we expect methods based on simple linear trends to perform quite well. In fact, in many countries all-cause mortality has been decreasing at a reasonably constant rate: we know, for example, that the Lee-Carter method performs well on U.S. data. (We also tried the Murray-Lopez model, but results were far less accurate.) When comparing our methods to the Lee-Carter model, there are some considerations to keep in mind, which play both in favor and against new approaches. The Lee-Carter method that we use in the comparison was developed to deal with U.S. data, and therefore it may not be fair to make comparisons with other data. On the other hand, we will be using our method in a very straightforward way, without making any

effort to obtain the best result for each country (e.g., we use the same specification for all countries).

We begin by studying a version of our approach by applying, country by country, the same smoothness functional of equation 11.12, which we report here:

$$H[\mu, \theta_{\text{age}}, \theta_{\text{time}}, \theta_{\text{age/time}}] \equiv \frac{\theta_{\text{age}}}{AT} \int_0^T dt \int_0^A da \left(\frac{d}{da}(\mu(a,t) - \bar{\mu}(a)) \right)^2 + \frac{\theta_{\text{time}}}{AT} \int_0^T dt \int_0^A da \left(\frac{d^2}{dt^2}\mu(a,t) \right)^2 + \frac{\theta_{\text{age/time}}}{TA} \int_0^T dt \int_0^A da \left(\frac{\partial^3 \mu(a,t)}{\partial a \partial t^2} \right)^2. \quad (12.1)$$

We do not smooth over countries using the Gibbs algorithm because that takes a long time, especially if we want to test the sensitivity of the results to the smoothness parameters. While this strategy is feasible, the long processing time would make life difficult for those readers who wish to reproduce our results. Instead, in order to retain at least some of the power of smoothing over countries, we take a slightly different approach: we use the preceding smoothness functional, obtain a set of forecasts, and then smooth the forecasts over countries, using generalized cross validation to choose an optimal smoothness parameter. This procedure is clearly suboptimal, but it is fast and will give us an idea of what smoothing over countries does.

Another way to measure the usefulness of any method is to measure not the goodness of the results of the method itself but of that method in conjunction with others. In recent years, it has become increasingly common to combine different methods, for example, by averaging or taking convex combinations of their forecasts. A natural choice here is to assess the performance of forecasts based on the average of the Lee-Carter forecast and of our method. It is reasonable to expect that the combined method will perform better than each of the two methods separately, because the two methods are quite different, and therefore the forecasts should not be very highly correlated.

In the end we decided to report the results of country-by-country analysis using five methods: Lee-Carter, the Bayesian method without smoothing over countries (Bayes 1), the Bayesian method with postprocessed smoothing over countries (Bayes 2), the average of Lee-Carter with Bayesian method without smoothing over countries (average 1), average of Lee-Carter and Bayesian method with postprocessing smoothing over countries (average 2). We measure the forecast error within one country as the mean absolute error averaged over all the age groups. We report results in terms of the percentage reduction in error (or percentage improvement in performance) over the Lee-Carter method applied alone.

We use the following specification for the purpose of this experiment:

$$\mu_{at} = \beta_a^{(0)} + \beta_a^{(1)} t + \beta_a^{(2)} \log(t - 1876) + \beta_a^{(3)} \log(GDP_t) + \beta_a^{(4)} \log(TOBACCO_{at}).$$

The tobacco variable is available for all countries except Hungary and Malta, and we have not attempted to find a close substitute for the purpose of this exercise, although it might be quite relevant for Hungary. The results for the 48 countries with more than 25 observations are reported in table 12.1. Across the four approaches, the median improvement over Lee-Carter is between 7.3% (with Bayes 1) and 14.6% (with Average 1). In general and as

TABLE 12.1.
Percentage Improvement of Four Methods over the Lee-Carter Method for all Causes in Males.

Country	*Bayes 1*	*Bayes 2*	*Average 1*	*Average 2*
Mauritius	−23.5	−31.4	−10.1	−13.8
Argentina	−65.9	−94.0	3.2	−6.2
Barbados	−25.2	−17.4	−8.6	−4.4
Belize	16.0	27.1	14.3	21.1
Canada	45.1	36.9	28.5	24.2
Chile	3.8	24.1	26.1	42.1
Colombia	0.8	−16.8	1.6	−5.5
Costa Rica	−108.6	9.9	−48.1	11.8
Cuba	−15.1	−9.4	0.9	5.0
El Salvador	−70.9	−9.8	−9.6	12.8
Mexico	28.6	18.9	34.6	45.9
Nicaragua	−53.6	−54.4	−19.1	−18.3
Panama	−130.0	−62.8	−36.5	−6.1
Suriname	28.1	26.2	15.7	15.1
Trinidad and Tobago	−4.4	0.7	20.7	14.4
USA	13.0	−6.5	27.9	11.8
Uruguay	−73.6	−90.7	−17.0	−17.1
Venezuela	1.5	−12.9	11.6	2.0
Israel	28.8	37.7	15.4	22.0
Japan	7.8	33.6	13.7	33.0
Kuwait	−1.4	−1.4	0.7	0.7
Singapore	17.9	20.8	13.3	13.6
Sri Lanka	−16.3	−9.9	−3.2	−3.4
Thailand	−26.8	−21.9	−7.9	−6.0
Austria	33.6	31.7	29.8	27.7
Belgium	12.6	22.4	12.1	16.9
Bulgaria	−0.2	−24.1	1.2	−11.7
Denmark	16.7	37.8	15.5	27.9
Finland	24.5	25.4	25.6	30.6
France	17.3	19.6	15.5	16.8
Germany	3.5	28.3	17.3	24.3
Greece	7.5	11.9	11.9	11.8
Hungary	−77.6	−40.8	−34.1	−19.0
Iceland	30.0	27.9	15.8	12.7
Ireland	43.6	62.2	26.3	34.7
Italy	32.4	41.2	23.6	27.0
Luxembourg	13.2	15.1	7.4	9.5
Malta	−1.9	14.0	8.7	16.4
Netherlands	20.1	28.4	19.9	17.9
Norway	42.0	49.1	20.8	24.5
Poland	−56.1	−35.7	−24.0	−15.2
Portugal	−25.3	−24.5	9.1	9.2
Spain	1.9	11.6	15.8	11.8
Sweden	30.5	7.1	14.5	3.6
Switzerland	20.1	22.9	18.4	14.1
United Kingdom	58.3	27.0	37.2	36.2
Australia	20.9	16.5	15.2	8.9
New Zealand	7.2	36.6	16.2	28.3
Median	7.3	14.6	14.0	12.8
25th quantile	−18.1	−13.9	0.9	−0.4
75th quantile	21.8	27.3	18.8	24.2

expected, the averages reduce the number of especially bad results, with the 25th quantile reduced from -18% or -13% to about zero.

12.2 Lung Disease in Males

We choose lung disease because it is a case in which we do have an interesting covariate, namely tobacco consumption. Mortality from lung disease has not been following a linear temporal pattern in most countries, and therefore we would not think that the Lee-Carter method would perform well, because the linear time trajectories are too constrained. The alternative is to use least squares, with tobacco as one of the covariates. In that case, however, we would expect that the performance will not be good for the opposite problem: too much variation in the time trajectories. We will describe the comparison with least squares later in the section. We use the same specification as in section 12.1.

In this example we introduce an alternative methodology for comparison. We still use the smoothness functional of equation 12.1 without smoothing over countries but take the approach of a user who wishes to analyze many countries at once without necessarily having to find the optimal smoothness parameters for each country. The question here is whether it would be difficult to find a single set of smoothness parameters that works in most cases. Another way to pose this question is to ask what would happen if a user picked the smoothness parameters at random in a wide interval: how often would our method outperform the base method? Or is there only a narrow range of parameters that guarantees better performance?

In order to answer these questions, we perform the comparison over a wide range of the parameters, and analyze the entire distribution of results. We can then, a posteriori, see what would have happened if an average user had made a reasonable choice of parameters. The range we have chosen for the parameters is very large, sampling 13 points on a logarithmic scale from 0.05 to 100.

The results of the experiment are displayed in a table with four columns: the first three columns are the smoothness parameters and the fourth column is the percentage improvement in mean absolute error over the Lee-Carter method. There are 2,197 rows in the table, corresponding to all 13^3 the possible combinations of the parameters. We are interested in finding out for which values of the parameters our method performs better than the Lee-Carter method. Ideally, we would plot the percentage improvement over Lee-Carter as a function of the three smoothness parameters. Because this is not possible in three dimensions, we "stratify" the plot: we plot the percentage improvement as a function of σ_{age} and $\sigma_{\text{age/time}}$ for several fixed values of σ_{time}. We choose σ_{time} as the stratifying variable because, as we demonstrate shortly, it explains most of the results in this particular data set. In figures 12.1 and 12.2, we plot the percentage improvement of our method over the Lee-Carter method as a function of the *logarithm* of σ_{age} and $\sigma_{\text{age/time}}$. The vertical axes of both Three-dimensional figures are the percentage improvement over Lee-Carter.

The first thing to notice in these two figures is that if σ_{time} is "too small" (meaning $\sigma_{\text{time}} \leq 0.2$), then our method always performs worse than the Lee-Carter method, independently of the values of σ_{age} and $\sigma_{\text{age/time}}$. However, when σ_{time} is equal to or larger than 0.6, then our method performs better than the Lee-Carter method, *independently of the values of* σ_{age} *and* $\sigma_{age/time}$. This implies that there is a huge range of values of the smoothness parameter for which our method performs better than Lee-Carter.

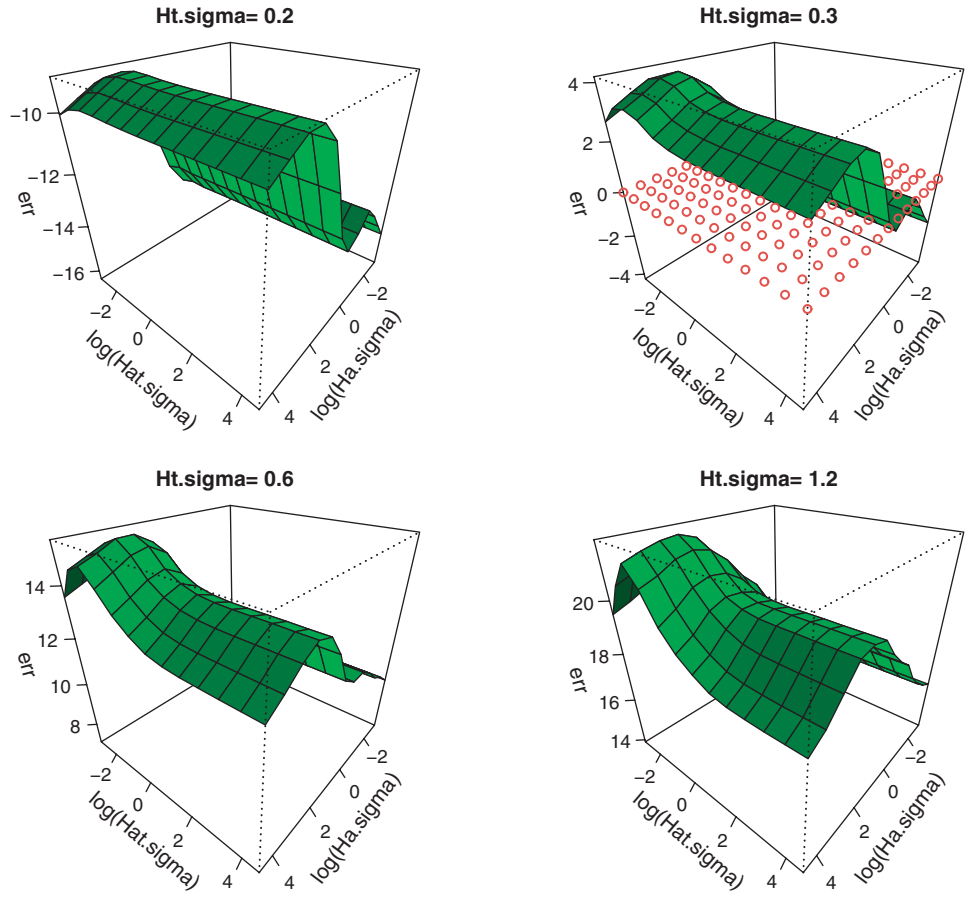

FIGURE 12.1. Percentage improvement of the MAP method over the Lee-Carter method plotted as a function of $\log(\sigma_{\text{age}})$ and $\log(\sigma_{\text{age/time}})$ for four different values of σ_{time}. The time series being forecast is lung disease in males age 30 or older. The red dots correspond to a percentage improvement equal to 0.

In order to better quantify the improvement over the Lee-Carter method, we summarize the information of figures 12.1 and 12.2 in table 12.2. In the first column of the table we report a value for σ_{time}, while in the second, third, and fourth columns we report the minimum, maximum, and median values of the percentage improvement, as σ_{age} and $\sigma_{\text{age/time}}$ vary in the interval [0.05, 100]. Table 12.2 makes clear that, because σ_{time} crosses some value between 0.33 and 0.63 (let us say 0.5 for simplicity), then the Bayesian method is always better than the Lee-Carter method, even if we pick arbitrory values of σ_{age} and $\sigma_{\text{age/time}}$ and use the same values for all the countries.

Thus, can we assure ourselves that any reasonable researcher would pick values of the smoothness parameters in the ranges where our method would improve the results? (The range for σ_{age} and $\sigma_{\text{age/time}}$ is so broad that there is no question about these parameters being set in a range where our approach would be advantageous.) The restriction that $\sigma_{\text{time}} \geq 0.5$ is more delicate, because values around 0.5 are not uncommon for smoothness parameters across applications. In order to show that it would be unusual to choose such a smoothness

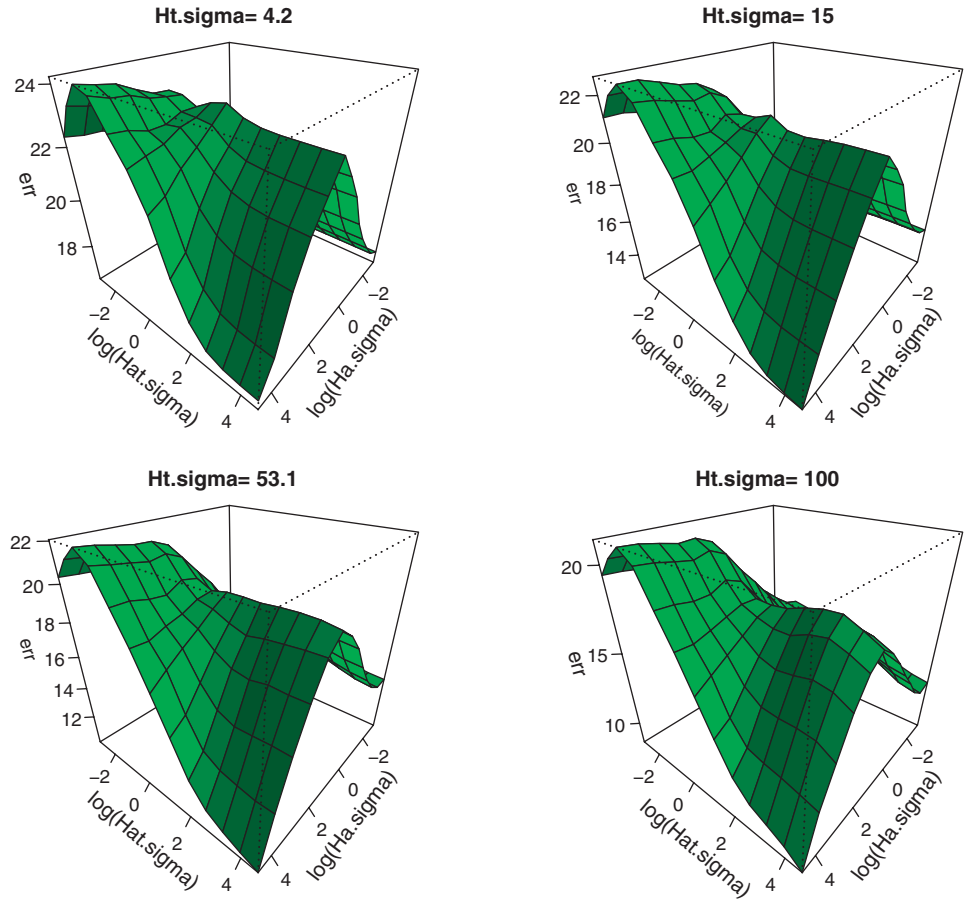

FIGURE 12.2. Percentage improvement of the MAP method over the Lee-Carter method plotted as a function of $\log(\sigma_{\text{age}})$ and $\log(\sigma_{\text{age/time}})$ for four different values of σ_{time}. The time series being forecast is lung disease in males age 30 or older.

parameter, we performed experiments on single countries, using our method for choosing the optimal choice of the smoothness parameters. In all cases we tested, the optimal value of σ_{time} was larger than 1.2, confirming that a value of 0.5 is neither reasonable nor likely.

12.2.1 Comparison with Least Squares

Once the comparison with the Lee-Carter method has been performed, it is easy to modify it in order to obtain a comparison with least squares. In fact, let e^{MAP}, e^{LC}, and e^{LS} be the median absolute error of our method, of the Lee-Carter (LC) method, and of the least-squares (LS) method respectively. The percentage improvement of our method over the Lee-Carter method is defined as

$$\Delta^{\text{MAP/LC}} \equiv \frac{e^{\text{LC}} - e^{\text{MAP}}}{e^{\text{LC}}}.$$

TABLE 12.2.
Percentage Improvement over the Lee-Carter, as a Function of σ_{time}.

σ_{time}	*min*	*max*	*median*
0.05	−29.58	−23.79	−27.81
0.09	−25.00	−18.81	−22.78
0.18	−16.38	−8.98	−13.02
0.33	−4.26	4.18	0.24
0.63	7.03	15.51	11.80
1.19	13.89	21.95	18.82
2.24	17.26	24.41	21.79
4.21	16.39	24.21	20.96
7.94	14.08	23.38	19.20
14.95	12.42	22.72	18.22
28.17	11.26	22.45	17.65
53.08	10.07	22.06	17.15
100.00	8.46	21.22	15.77

Note: For a given value of σ_{time} the values of σ_{age} amd $\sigma_{\text{age/time}}$ vary between 0.05 and 100, and the resulting range in percentage improvement is reported.

Similarly we define

$$\Delta^{\text{MAP/LS}} \equiv \frac{e^{\text{LS}} - e^{\text{MAP}}}{e^{\text{LS}}}.$$

and

$$\Delta^{\text{LC/LS}} \equiv \frac{e^{\text{LS}} - e^{\text{LC}}}{e^{\text{LS}}}.$$

Some simple algebra shows that the following relationship holds:

$$\Delta^{\text{MAP/LS}} = \Delta^{\text{MAP/LC}} + \Delta^{\text{LC/LS}} - \Delta^{\text{LC/LS}}\Delta^{\text{MAP/LC}}.$$

The important term in this sum is the improvement of the Lee-Carter method with respect to least squares, $\Delta^{\text{LC/LS}}$, which is equal to 11.9%. The last term is a second-order concern, which is at most 2%. Therefore, overall, the percentage improvement of our method over least squares is about 10 percentage points more than the percentage improvement over the Lee-Carter method.

12.2.2 Country-by-Country Analysis

We also performed an analysis similar to the one performed for all cause mortality, in which we estimated country-specific parameters for 48 countries. Results for this analysis are shown in table 12.3. Although there are several countries for which our method does not perform well, the median improvement of the Bayesian method is a respectable 21.5%, which climbs to 24.9% when some country smoothing is imposed. Averaging the Bayesian and the Lee-Carter model does not lead to any improvement in this case.

TABLE 12.3.
Percentage Improvement of Four Methods over the Lee-Carter Method for Lung Disease in Males.

Country	*Bayes 1*	*Bayes 2*	*Average 1*	*Average 2*
Mauritius	2.6	1.9	6.8	5.2
Argentina	23.6	−105.4	16.1	−20.1
Barbados	23.7	34.4	20.0	30.2
Belize	22.1	24.8	18.3	37.3
Canada	46.0	70.0	26.1	53.6
Chile	14.0	−36.1	26.15	10.7
Colombia	4.7	−33.7	25.2	11.8
Costa Rica	−20.2	−37.5	−10.2	−17.1
Cuba	20.9	42.3	37.2	50.4
El Salvador	11.1	15.2	13.5	12.0
Mexico	25.6	−30.4	22.4	−8.3
Nicaragua	27.3	34.7	30.4	24.8
Panama	−122.0	−72.3	−57.3	−32.5
Suriname	45.0	47.1	30.8	33.9
Trinidad and Tobago	1.7	−6.5	3.6	1.5
USA	56.8	54.9	44.8	66.8
Uruguay	−83.0	26.7	−27.0	30.3
Venezuela	4.6	11.7	26.9	28.1
Israel	33.5	31.9	21.4	21.3
Japan	14.3	33.1	48.0	49.1
Kuwait	−7.1	−7.1	−0.7	−0.7
Singapore	25.3	39.4	12.9	20.5
Sri Lanka	−37.4	−27.2	−14.5	−9.3
Thailand	−11.9	−7.5	10.5	13.0
Austria	17.6	24.9	37.2	30.2
Belgium	24.7	39.9	17.1	49.6
Bulgaria	−41.5	−31.0	1.9	1.9
Denmark	29.0	32.8	15.5	24.9
Finland	20.6	45.7	23.4	30.8
France	36.6	20.1	26.1	22.7
Germany	−20.7	19.0	−9.1	8.3
Greece	14.4	31.3	12.89	18.9
Hungary	−63.0	−17.7	−5.2	15.6
Iceland	−1.3	−12.7	10.2	8.6
Ireland	52.6	65.8	41.8	35.1
Italy	49.8	64.1	25.3	33.3
Luxembourg	10.8	11.5	20.0	25.7
Malta	−4.3	6.6	9.3	14.5
Netherlands	59.2	79.0	37.5	50.0
Norway	35.5	27.8	20.1	15.6
Poland	25.6	24.9	19.0	21.0
Portugal	−2.2	−30.4	2.2	−8.7
Spain	31.9	34.5	16.9	26.1
Sweden	38.9	−23.0	28.9	−10.4
Switzerland	31.0	35.7	13.2	20.4
United Kingdom	66.2	77.9	58.3	58.3
Australia	70.6	52.6	42.2	28.4
New Zealand	55.5	52.4	37.8	27.4
Median	21.5	24.9	19.5	21.2
25th quantile	0.9	−8.8	10.0	8.5
75th quantile	34.0	39.5	27.4	30.4

12.3 Breast Cancer in Females

Next we study breast cancer in females. In this case we do not have a good set of covariates, that is, covariates known to reliably predict breast cancer. Instead, we choose a specification similar to the one for lung disease, to which we add human capital. The specification is as follows:

$$\begin{aligned}\mu_{at} =& \beta_a^{(0)} + \beta_a^{(1)} t + \beta_a^{(2)} \log(t - 1876) \\ &+ \beta_a^{(3)} \log(GDP_t) + \beta_a^{(4)} \log(TOBACCO_{at}) + \beta_a^{(5)} \log(HC_t).\end{aligned}$$

It would be surprising if we found levels of improvements similar to the ones observed for lung disease, because the covariates are not optimal. We perform the experiment and report the results in the same way we did for lung disease. The percentage improvement of our method over the Lee-Carter method is shown in figure 12.3 for different values of σ_{time}.

Unlike in the case of lung disease, here the range of values of σ_{time} that produce improvement is not extremely large: for $\sigma_{\text{time}} > 7$, our method performs worse than Lee-Carter. However, small values of σ_{time} are those which produce the best results: for $\sigma_{\text{time}} = 0.59$, our method performs better than Lee-Carter for *any* values of σ_{age} and $\sigma_{\text{age/time}}$, with improvement up to 11%. The question here is whether any researcher would set the values of the smoothness parameters in intervals that guarantee good results. As in the case of lung disease, we use our method for determining the optimal values of the smoothness parameters on several countries and find optimal values in the following range:

$$\sigma_{\text{age}} \in [0.05, 1.7] \,, \quad \sigma_{\text{age/time}} \in [1, 4] \,, \quad \sigma_{\text{age}} \in [1, 2].$$

Using values in these ranges, we find that our method always performs better than Lee-Carter, with a median improvement of 9.6%. This results support the notion that, setting the smoothness parameters in "reasonable" ranges, one would always obtain results that are better than Lee-Carter.

12.3.1 Comparison with Least Squares

Because the specification we use has four covariates, we expect the least-squares method to perform very poorly and exhibit a huge variance. This indeed the case: we find that the improvement of Lee-Carter over least squares is 61%. Thus, even when our method performs worse than Lee-Carter, it performs better than least squares. Another way to state this result is that even the wrong amount of smoothing is preferable to no smoothing at all.

12.3.2 Country-by-Country Analysis

We also performed an analysis similar to the one performed for all-cause mortality, in which we estimated country-specific parameters for 48 countries. The results, shown in table 12.4, are not as good as those for lung cancer, although they are better than those obtained for all causes, with a median improvement of 13.2%, which climbs to 17.6% when country smoothing is introduced. As in the lung disease case, averaging the Bayes and Lee-Carter method does not improve the results.

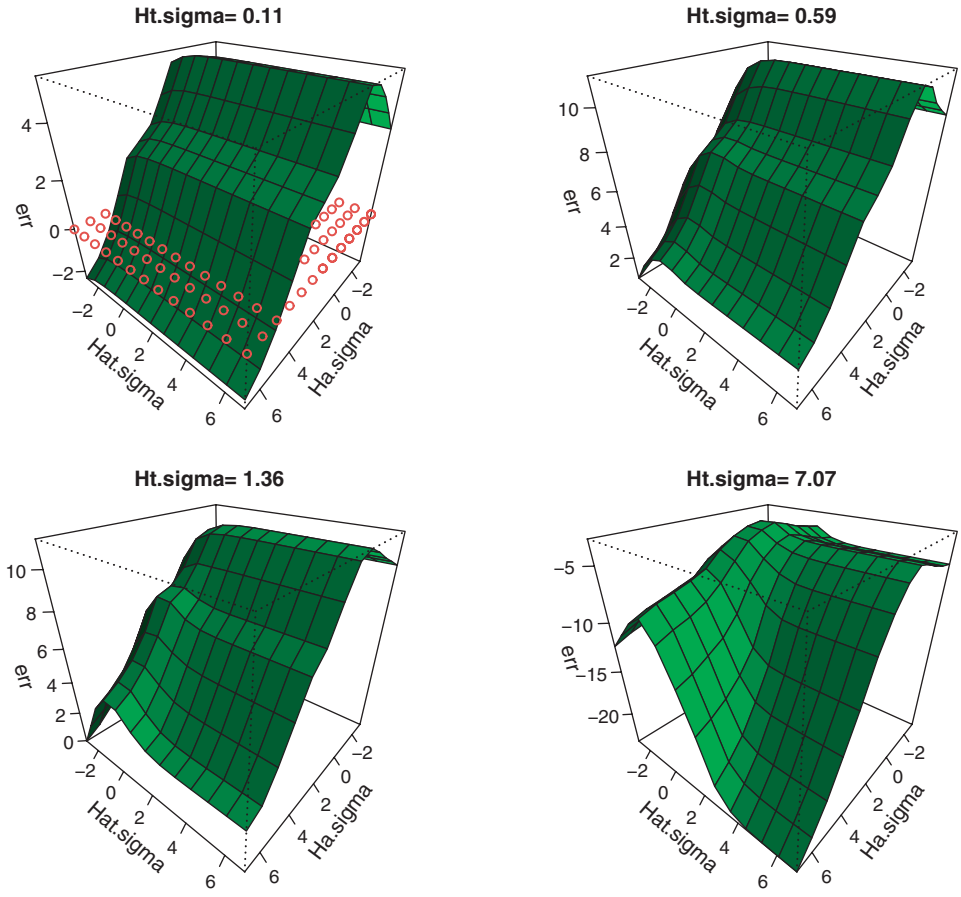

FIGURE 12.3. Percentage improvement of the MAP method over the Lee-Carter method for breast cancer in females, plotted as a function of $\log(\sigma_{\text{age}})$ and $\log(\sigma_{\text{age/time}})$ for four different values of σ_{time}. The red dots correspond to a percentage improvement equal to 0. They have been reported to make it easier to see that for very large values of $\log(\sigma_{\text{age}})$ and $\log(\sigma_{\text{age/time}})$ the performance of our method drops below the performance of the Lee-Carter method.

12.4 Comparison on OECD Countries

So far we have performed our comparisons on the same set of 48 countries, originally chosen as those whose time series for all-cause mortality has more than 25 observation in the period 1950–2000. It is natural to ask how the results change if we restrict that original set to some more specific and homogeneous set of countries. Data like these therefore should have more statistical strength that is available to borrow with our method from neighboring cross sections. Here we consider the case in which we select, out of those 48 countries, those that belong to the Organization for Economic Co-operation and Development (OECD).[1]

[1] In so doing, we do not represent the entire OECD: we lose five countries

TABLE 12.4.
Percentage Improvement of Four Methods over the Lee-Carter Method for Breast Cancer in Females.

Country	*Bayes 1*	*Bayes2*	*Average 1*	*Average 2*
Mauritius	21.5	27.4	10.8	13.0
Argentina	21.3	−30.6	10.7	−14.3
Barbados	45.6	46.2	31.7	31.7
Belize	29.6	40.9	14.8	20.9
Canada	44.2	46.3	29.9	26.8
Chile	14.5	−27.6	11.2	−10.7
Colombia	−4.8	20.4	16.6	26.7
Costa Rica	13.5	28.2	21.1	28.0
Cuba	−157.6	−86.9	−56.8	−22.9
El Salvador	2.9	15.1	2.5	8.9
Mexico	28.9	−0.5	39.0	20.9
Nicaragua	−70.5	−39.1	−25.8	−12.7
Panama	2.0	12.7	7.0	7.5
Suriname	44.3	44.6	34.0	33.3
Trinidad and Tobago	17.3	21.7	10.4	10.6
USA	24.4	43.6	20.1	27.9
Uruguay	28.5	42.2	19.6	28.9
Venezuela	13.0	44.6	9.4	31.7
Israel	34.8	31.5	20.6	19.4
Japan	23.4	−42.2	39.1	55.3
Kuwait	32.1	32.1	25.9	25.9
Singapore	25.9	32.3	19.8	21.4
Sri Lanka	−18.7	−16.8	−9.0	−7.9
Thailand	24.5	23.4	12.9	12.4
Austria	−20.0	−20.4	5.8	3.4
Belgium	3.8	36.0	14.4	26.3
Bulgaria	7.7	3.5	10.4	11.9
Denmark	18.8	11.2	15.9	6.2
Finland	36.9	26.4	29.6	38.5
France	3.5	−4.2	6.7	−1.8
Germany	0.8	19.3	2.5	10.5
Greece	45.9	37.2	28.0	25.3
Hungary	−5.3	3.4	−1.6	2.0
Iceland	6.6	13.8	7.2	8.8
Ireland	8.6	22.3	17.5	21.4
Italy	1.1	16.0	9.2	17.1
Luxembourg	27.4	32.8	19.1	19.4
Malta	−10.9	−15.5	12.9	10.8
Netherlands	26.0	13.0	24.2	21.8
Norway	0.9	15.0	8.4	14.5
Poland	21.3	38.9	28.2	26.7
Portugal	−9.4	6.2	3.9	11.1
Spain	49.0	46.5	25.3	27.8
Sweden	−1.3	−10.9	11.7	1.5
Switzerland	−17.7	7.2	2.1	18.0
United Kingdom	−15.6	49.8	1.5	32.0
Australia	−46.6	−31.9	−11.5	−4.7
New Zealand	−41.8	−20.7	−6.8	5.3
Median	13.2	17.6	12.3	17.5
25th quantile	−2.2	−1.4	6.5	7.2
75th quantile	26.3	33.6	20.7	26.7

TABLE 12.5.
Percentage Improvement of Four Methods over the Lee-Carter Method for Transportation Accidents in Males: Most OECD Countries.

Country	*Bayes 1*	*Bayes 2*	*Average 1*	*Average 2*
Canada	44.7	50.2	44.2	37.1
USA	7.5	−51.2	14.4	8.7
Japan	−172.6	−133.3	−23.4	−17.0
Austria	46.5	46.0	52.5	55.2
Belgium	−46.2	−77.8	32.2	21.0
Bulgaria	71.4	72.2	51.9	46.7
Denmark	−50.1	−35.6	19.8	24.6
Finland	4.0	5.2	56.9	57.0
France	18.5	27.9	43.1	44.0
Germany	22.4	31.3	14.4	19.1
Greece	37.4	12.2	56.8	61.8
Hungary	59.8	68.4	32.2	37.8
Iceland	17.1	24.3	8.9	13.4
Ireland	38.8	45.2	19.9	25.7
Italy	−65.6	−38.9	5.8	11.2
Netherlands	−257.5	−72.5	−62.6	−3.4
Norway	44.2	47.0	36.0	36.8
Poland	73.5	68.1	44.3	50.3
Portugal	57.6	70.3	30.1	42.3
Spain	68.4	75.6	52.4	52.7
Sweden	27.8	30.4	24.4	21.2
Switzerland	25.5	31.1	27.8	30.9
United Kingdom	63.6	37.0	44.7	22.9
Australia	34.1	45.6	31.5	35.9
New Zealand	5.6	29.0	6.0	23.1
Median	27.8	31.1	31.5	30.9
25th quantile	5.6	5.2	14.4	21.0
75th quantile	46.5	47.0	44.3	44.0

12.4.1 Transportation Accidents in Males

We do not have very good covariates for predicting transportation accidents. We do not expect the time series to follow a linear pattern, so we certainly want a nonlinear trend in our specification. In terms of other covariates, it seems reasonable to include *GDP*, because economic growth can certainly influence the amount of traffic on the roads, and therefore the number of accidents. We would then need a variable that controls for the increased safety of the roads, which captures at least some aspects of public health policy. For the lack of a better candidate, we chose tobacco consumption, which is a variable that is affected by safety and public health concerns: one could argue that countries with more rapid decline of tobacco consumption are also those countries where there is more activity in the public health arena, and therefore those which are more likely to implement road safety measures.

Clearly, covariates that predict transportation accidents are available, and so it should be possible to greatly improve these forecasts and the extent to which our method

TABLE 12.6.
Percentage Improvement of Four Methods over the Lee-Carter Method for Cardiovascular Disease in Males.

Country	*Bayes 1*	*Bayes 2*	*Average 1*	*Average 2*
Canada	24.8	55.6	21.8	42.2
USA	−8.2	−10.4	23.7	28.8
Japan	14.1	49.9	51.0	52.6
Austria	58.2	64.1	32.5	36.1
Belgium	46.1	44.4	35.3	29.2
Bulgaria	30.0	−46.0	17.4	−17.7
Denmark	26.6	21.9	24.2	21.6
Finland	54.5	63.6	36.2	43.2
France	−14.0	−39.6	27.5	11.5
Germany	44.8	57.8	23.1	32.8
Greece	−12.4	62.4	−3.1	−13.0
Hungary	−69.7	−42.0	−27.0	−14.5
Iceland	44.8	8.4	31.2	14.2
Ireland	38.1	57.6	28.5	41.1
Italy	75.3	73.2	48.9	39.8
Netherlands	64.0	62.7	36.8	38.0
Norway	53.1	44.9	38.8	36.0
Poland	−40.0	−16.9	−12.2	0.3
Portugal	26.9	33.0	26.8	33.2
Spain	−0.9	34.8	15.4	24.1
Sweden	33.9	52.1	27.9	29.3
Switzerland	55.1	52.5	36.5	32.9
United Kingdom	25.4	65.2	15.8	38.2
Australia	26.3	38.2	23.8	30.7
New Zealand	39.2	40.5	28.5	31.7
Median	30.0	44.4	27.5	31.7
25th quantile	14.1	8.4	21.8	21.6
75th quantile	46.1	57.6	35.3	38.0

outdistances others. For our methodological purposes, we have not attempted to collect them. Our final specification is therefore:

$$\mu_{at} = \beta_a^{(0)} + \beta_a^{(1)} t + \beta_a^{(2)} \log(t - 1876) + \beta_a^{(3)} \log(GDP_t) + \beta_a^{(4)} \log(TOBACCO_{at}).$$

The results are shown in table 12.5. Despite some notable failures, such as Japan, the median improvement is of the order of 30%. If one considers that the results were obtained literally by "pushing a button," that no cross validation was involved, and that we invested no effort in finding the right covariates, this seems encouraging. Part of the success is due to the fact that we restricted ourselves to a smaller and more homogeneous set of countries. In fact, had we started from the original set of 48 countries, the average improvement would have been 18% without country smoothing and 25% with country smoothing. In the final set of examples, we consider a case where a more appropriate set of covariates is available.

12.4.2 Cardiovascular Disease in Males

For cardiovascular disease we have a better choice of covariates: tobacco consumption and fat consumption. Unlike the tobacco consumption covariate, which is age-specific, the fat consumption covariate does not vary with age, and so the specification we use is as follows:

$$\begin{aligned}\mu_{at} =& \beta_a^{(0)} + \beta_a^{(1)} t + \beta_a^{(2)} \log(t - 1876) \\ &+ \beta_a^{(3)} \log(GDP_t) + \beta_a^{(4)} \log(TOBACCO_{at}) + \beta_a^{(5)} \log(FAT_t).\end{aligned}$$

In addition, while we have tobacco consumption for all but one OECD countr y, the fat consumption covariate is missing in 25% of the countries we analyze. Despite these shortcomings, we still expect reasonably good results.

The results are shown in table 12.6 and are indeed quite favorable: the improvement over Lee-Carter is 30% or 44% depending on whether we use or do not use country smoothing. Part of the success is because we limited ourselves to a set of more homogeneous countries with better covariates. If we perform the same analysis on the original 48 countries, however, the fat covariate would be missing in 50% of the cases, but the median improvement would still be about 21.5%.