

# **Bayesian Forecasting of Cohort Fertility**

Carl Schmertmann    *Florida State University*  
Emilio Zagheni      *Queens College, City University of New York*  
Joshua Goldstein    *University of California, Berkeley*  
Mikko Myrskylä      *London School of Economics*

## Introduction

Fertility has experienced long-term declines in many developed countries, but recent studies have noted small increases in some parts of Europe, North America, and Asia (Goldstein, Sobotka and Jasilioniene 2009; Luci and Thevenon 2010; Myrskylä, Kohler and Billari 2009; Sobotka 2008; Myrskylä, Goldstein and Chen 2013). The future size and age structure of national populations depend largely on birth rates, so governments and planners are naturally interested in knowing whether fertility decline is slowing or reversing.

Answering that question with current data is complicated. Statistical agencies estimate the most common fertility index – the *total fertility rate* (TFR) – by aggregating the age-specific fertility rates of a calendar year. Thus a TFR such as 1.16 children per woman (for the Czech Republic in 1998) does not correspond to average lifetime childbearing by any real women, but rather to a fictitious group who experience 1998 age-specific rates over their reproductive lifetimes. This kind of period measure is vulnerable to what demographers call *tempo distortion*. In particular, an increase in TFR does not necessarily mean that women are beginning to have larger families. It could instead mean that postponement of fertility to higher maternal ages is slowing (Bongaarts and Feeney 1998; Van Imhoff and Keilman 2000; Kohler and Philipov 2001; Zeng and Land 2002; Goldstein et al. 2009).

Simply put, standard fertility indices derived from current annual birth and population information cannot tell us the future. Estimating the final fertility of *cohorts* (i.e., real groups of women, such as those born in 1970, 1980, or 1990) requires either waiting for those women to reach the end of reproductive ages, or making forecasts. The waiting strategy has produced a few recent examples of increased cohort fertility in Scandinavia (Andersson et al. 2009), but in most

countries the cohorts of women whose lifetime fertility might be increasing are still young, and their final fertility levels depend on future rates (Frejka and Calot 2001).

Our objective in this paper is to produce useful forecasts of completed cohort fertility for women born in the 1970s and 1980s, by using a combination of new methods and a large new fertility database. Figure 1 shows an example forecasting problem, using data from the Czech Republic<sup>1</sup>. Precise rate estimates are available for Czech women by single years of age for calendar years through 2009 from the Human Fertility Database (HFD 2011, described in detail later). This yields complete rate histories over ages 15-44 for women born 1956-1965, and partial histories for cohorts of women born after 1965. The data indicate a radical change in Czech fertility patterns in the post-Communist era. The top edge of the figure contains values for the *cohort total fertility rate* (CFR), which is the average number of children ever born to women with a particular birth year. The previous Czech pattern of early births and replacement-level cohort fertility (CFR $\approx$ 2.1) has given way to later births and a still-unknown level of completed fertility. The central question that a forecast must address is whether women in the later cohorts are reducing fertility, or merely postponing it. The same question arises in other developed countries.

---

<sup>1</sup> In our data the Czech Republic and Slovakia are separated, even for the period during which they were politically united.

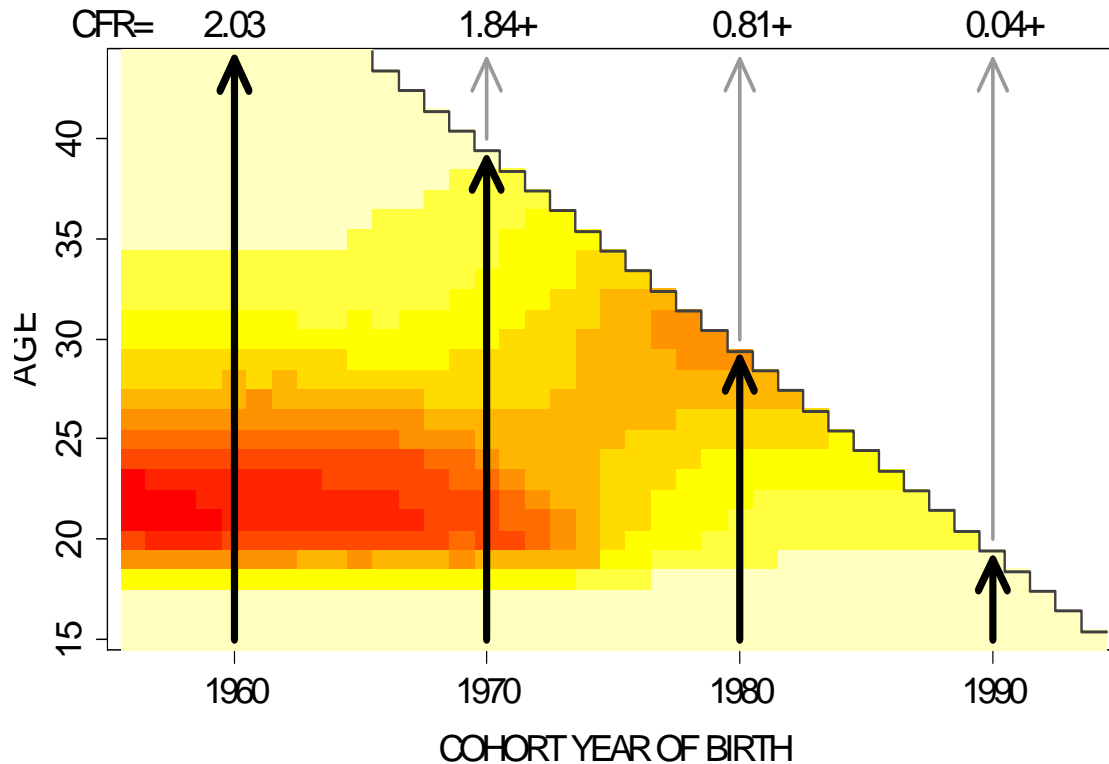


Figure 1. Czech Republic HFD (2011) fertility estimates by cohort year of birth and single year of age, for women born after 1955. Darker cells indicate higher rates. Cells in the upper right, such as (born in 1980, age 40), will occur in the future. Cohort fertility rates (CFR) appear above the plot, and are incomplete for cohorts born after 1965.

## Demographic Forecasting

Recent changes in fertility levels and timing make our forecasting question especially timely, but of course the basic problem is not new to demography. There is a growing literature in forecasting mortality, much of it derived from Lee and Carter's (1992) singular value decomposition approach for period mortality forecasts (e.g., Renshaw and Haberman 2006; Booth and Tickle 2008; Girosi and King 2008; Hyndman and Booth 2008). Girosi and King's (2008) recent proposals for Bayesian models in mortality forecasting are especially important: they are a main inspiration for the cohort fertility models that we develop in this paper.

Fertility forecasting is a far more difficult problem. Unlike death, childbearing is both optional and repeatable. Its timing is strongly affected by conscious decisions. In addition, mortality rates change predictably in one direction over time, while fertility rates fluctuate. Despite these difficulties, there is a sophisticated literature, parallel to that for mortality, on forecasting period fertility rates and the completed fertility of cohorts (Bloom 1982; de Beer 1985; Thompson et al. 1989; Chen and Morgan 1991; Lee 1993; Li and Wu 2003; Goldstein 2008; Hyndman and Booth 2008; Chen 2010; Cheng and Lin 2010; Alkema et al. 2011; Myrskylä, Goldstein and Chen 2013). Booth (2006) offers an excellent overview and history of demographic forecasting in general, including a discussion of approaches to cohort fertility completion.

Models for cohort fertility forecasts can extrapolate over time, over ages, or both. In Figure 1, a time series approach would extrapolate available age-specific rates from West to East, while a cohort approach would fit models to data for incomplete cohorts and extrapolate from South to North over ages. A principal challenge for any model is to produce coherent forecasts, in which both time trends and cohort schedules are demographically plausible. In particular, it is essential to use recent trends (particularly on postponement to higher maternal ages) in a way that avoids forecasts with demographically implausible age patterns for cohorts.

Another important forecasting challenge is the evaluation of uncertainty. Deterministic models that employ simple cohort extrapolations or parametric trajectories for future age-specific rates do not come with straightforward and well-developed methods for evaluating forecast uncertainty.

In this paper we extrapolate over both time and age, combining what demographers already know about plausible age patterns of fertility with recent trends in age-specific rates. We

build a Bayesian model for surfaces such as those in Figure 1, with priors constructed from a large archive of historical fertility data. As described in the next sections, our model uses improper normal priors and a normal likelihood, so that calculation of posterior maxima and posterior uncertainty is straightforward.

Our approach uses available information without imposing rigid models or making strong assumptions about the object of the forecast. Our goal is to use Bayesian methods with priors that explicitly incorporate some of the knowledge that is implicitly used by existing forecasts methods. We expect that such methods may find better compromises between recent trends in observed data and known age, period, and cohort fertility patterns. In developing this model, our main tasks are to define qualitative priors that include existing demographic knowledge about fertility patterns over age-time Lexis surfaces, to devise appropriate mathematical specifications for those priors, and to design appropriate computational methods.

### **Human Fertility Database and Supplemental Data Sources**

We base our analysis on a new public dataset, the Human Fertility Database (HFD 2011). The HFD is a cooperative project of the Max Planck Institute for Demographic Research and the Vienna Institute of Demography. We downloaded data comprising 44400 estimated fertility rates for 24 countries or regions in Europe and North America, categorized by single-year of age (12,13,...,55) and single calendar year of data collection, over periods of up to 120 years ending near 2009.<sup>2</sup> HFD (2011) protocols carefully allocate these data to women by their year of birth.

---

<sup>2</sup> Our HFD period data are for Austria 1951-2008, Bulgaria 1947-2009, Canada 1921-2007, Switzerland 1932-2007, Czech Republic 1950-2009, Germany [East, West, combined] 1956-2010, Estonia 1959-2009, Finland 1939-2009, France 1946-2009, UK [N.Ireland 1974-2009, Scotland 1945-2009, England & Wales 1938-2009, combined 1974-2009], Hungary 1950-2009, Lithuania 1959-2009, Netherlands 1950-2009, Portugal 1940-2009, Russia 1959-2009, Slovakia 1950-2009, Slovenia 1983-2009, Sweden 1891-2010, USA 1933-2007.

We combined HFD data with period rates collected by Myrskylä and colleagues for 12 additional countries (Australia, Belgium, Denmark, Greece, Iceland, Italy, Japan, Korea, Luxembourg, New Zealand, Romania, and Singapore; data sources are listed in Myrskylä et al. 2013, Table 1), and also with period rates for Brazil 1966-2010 provided by Dr. Everton Lima (personal communication). This produced a final data set covering 37 countries, which is heavily European but also includes data from North America, South America, Asia, and Oceania.

In this data set, we call the vector of rates for a cohort at ages 15-44 its *cohort fertility schedule*, and define a schedule as *complete* if rate estimates are available at all 30 ages. As an example, in Figure 1 the schedule for Czech women born in 1960 is complete, while the schedule for women born in 1980 is not. We have data from 3223 cohort schedules, of which 1015 are complete. The earliest complete schedule is for Swedish women born in 1876; the latest are for women born in the mid-1960s in each country.

We separate our data into two non-overlapping subsets:

- *Contemporary* data for forecasting exercises over surfaces such as Figure 1.
- *Historical* data, for use as a source of *a priori* information. Historical data comprise all complete fertility histories for cohorts born earlier than those appearing in any forecast surface.

The historical data set contains  $S=469$  complete cohort schedules for women born in any country between 1900 and 1949.<sup>3</sup> We organize the historical dataset as a  $30 \times 469$  matrix **HIST**, with each column containing one complete historical schedule. Contemporary data for each country (such as that for the Czech Republic in Figure 1) comprise fertility schedules, both complete and incomplete, for all women born after 1949.

---

<sup>3</sup> Only Swedish data are available for cohorts born before 1906. Choosing 1900 as a lower bound ensures that priors come from a more equally weighted mix of countries. The upper bound of 1949 ensures that, as in Figure 1, we have enough complete cohorts to observe pre-forecast trends in rates by age.

The HFD represents an unprecedented collection of coherent and comparable fertility data. Combining the HFD with supplemental data creates a contemporary data set that is an excellent testbed for comparative forecasts. Furthermore, the historical portion of the data set is an ideal source of *a priori* information on the nature of cohort fertility schedules and their changes over time. The use of a large collection of historical data to inform and construct improper priors about age and time patterns of cohort fertility is one of the main novelties in our forecasting exercise.

### Notation, Model, and Methods

For contemporary data in a country, we have  $C$  birth cohorts of interest ( $c=1\dots C$ ) over  $A$  reproductive ages ( $a=1\dots A$ ). For this country, define (with all vectors as columns):

- $\theta_{ca} \in R$ , the true fertility rate for cohort  $c$  between exact ages  $a$  and  $a+1$
- $\theta_c = (\theta_{c1} \dots \theta_{cA})' \in R^A$ , the *fertility schedule* for cohort  $c$ ,
- $\theta_a = (\theta_{1a} \dots \theta_{Ca})' \in R^C$ , the *time series* of rates at age  $a$ ,
- $\theta = (\theta'_1 \dots \theta'_C)' \in R^{CA}$ , the vector of *all rates*, sorted by age within cohort,
- $\mathbf{G}_c = [\mathbf{0} \dots \mathbf{I}_A \dots \mathbf{0}] \in R^{A \times CA}$ , a matrix such that  $\theta_c = \mathbf{G}_c \theta$
- $\mathbf{H}_a = \mathbf{I}_C \otimes (0 \dots \underset{a\text{-th}}{1} \dots 0) \in R^{C \times CA}$ , a matrix such that  $\theta_a = \mathbf{H}_a \theta$
- $CFR_c = (1 \dots 1) \theta_c \in R$ , the *completed fertility* of cohort  $c$
- $y \in R^n$ , a vector of published estimates for some subset of  $\theta$
- $\mathbf{V} \in R^{n \times CA}$ , a matrix of ones and zeroes such that  $\mathbf{V} \theta \in R^n$  is the subset of parameters corresponding to  $y$



In the forecasting problem,  $y$  is a set of external estimates of past fertility rates, provided by a national statistical agency. As in Figure 1, the  $C \times A$  grid  $\theta$  (also called a *rate surface* or a *Lexis surface*) is a larger set that not only includes those past rates, but also extends into the future.

We model parameters  $\theta$  and observations  $y$  in a Bayesian framework

$$(1) \quad \ln P(\theta|y) = \text{const} + \ln L(y|\theta) + \ln f(\theta)$$

where  $P(\cdot)$ ,  $L(\cdot)$ , and  $f(\cdot)$  represent, respectively, the posterior density, the likelihood function, and the prior density.  $\text{Const}$  is a term that does not vary with  $\theta$ . Fertility rates published by national statistical agencies typically come from very large risk populations. This justifies a normal approximation for the likelihood,

$$(2) \quad \ln L(y|\theta) = \text{const} - \frac{1}{2} (y - \mathbf{V}\theta)' \mathbf{\Psi}^{-1} (y - \mathbf{V}\theta)$$

where  $\mathbf{\Psi} = \text{diag} [ y_i(1-y_i)/W_i ]$  and  $W_i$  is the number of  $a$ -year-old women in the  $(c,a)$  cell corresponding to the  $i$ -th rate. In most cases  $W_i$  values are very large, so that sampling variances are near zero and estimates  $y$  are almost always extremely close to the true fertility rates in the pre-forecast period.

As we describe in detail below, we use a log prior density of the form

$$(3) \quad \ln f(\theta) = \text{const} - \frac{1}{2} \theta' \mathbf{K} \theta$$

where the constants in  $CA \times CA$  matrix  $\mathbf{K}$  are estimated from patterns in historical data. In combination with the normal likelihood function (2), this prior implies, conditional on  $\mathbf{K}$ , a multivariate normal posterior for  $\theta|y$ , with  $CA \times 1$  mean vector

$$(4) \quad \mu_{\text{post}} = [\mathbf{V}' \mathbf{\Psi}^{-1} \mathbf{V} + \mathbf{K}]^{-1} [\mathbf{V}' \mathbf{\Psi}^{-1} y]$$

and  $CA \times CA$  covariance matrix

$$(5) \quad \Sigma_{\text{post}} = [\mathbf{V}' \mathbf{\Psi}^{-1} \mathbf{V} + \mathbf{K}]^{-1}$$

The posterior distribution over each country's Lexis surface,

$$(6) \quad (\boldsymbol{\theta} \mid y_{country}) \sim N(\boldsymbol{\mu}_{post}, \boldsymbol{\Sigma}_{post}),$$

serves as a probabilistic forecast for its future age-specific fertility rates. Because

$CFR_c = (1 \dots 1)\boldsymbol{\theta}_c = \mathbf{1}'\mathbf{G}_c \boldsymbol{\theta}$  is a linear function of  $\boldsymbol{\theta}$ , the posterior distribution also provides a probabilistic forecast for our primary measure of interest, completed cohort fertility.

The critical part of this model is the very large penalty matrix  $\mathbf{K}$  that specifies the improper prior. Our basic approach, similar to that in Girosi and King (2008) or Wood (2000), is to build  $\mathbf{K}$  from the bottom up, by additively combining many sub-penalties that apply to the individual cohort schedules and to the time series of fertility rates at each age. The process of repeated penalties builds a scaffolding for a partially-complete rate surface like that in Figure 1, with vertical (cohort) and horizontal (time-series) beams that extend from the past into the future.

We construct these sub-penalties by combining demographic knowledge with empirical patterns in historical fertility data. In particular, we assign lower *a priori* probabilities to historically implausible  $\boldsymbol{\theta}$  surfaces that have (1) age patterns in cohort fertility schedules  $\boldsymbol{\theta}_1 \dots \boldsymbol{\theta}_C$  that are unlike those in the historical data [via *cohort penalties*], and (2) patterns in time series of age-specific rates  $\theta_{15} \dots \theta_{44}$  that are unlike corresponding series in the historical data [*time series penalties*]. Details are in the next subsections.

### *Cohort Penalties*

In all of the examples and calculations for this paper, we deal with rate surfaces over exactly  $A=30$  ages (15...44) and  $C=40$  cohorts (women born in 1956...1995), and we will assume those values from this point forward. For each cohort schedule  $\boldsymbol{\theta}_c$  we define historically

unlikely age patterns via the singular value decomposition (SVD) of the historical data array,  $\mathbf{HIST}=\mathbf{UDV}'$ . Call  $\mathbf{X}$  the  $30 \times 3$  matrix constructed from the mutually orthogonal  $\mathbf{U}$  columns corresponding to the three largest singular values in  $\mathbf{D}$ . Figure 2 shows these three  $\mathbf{X}$  columns, which have clear demographic interpretations. Weights on components 1-3 affect the overall cohort fertility level, the mean age of childbearing, and the variance of childbearing ages, respectively. Increases in component 2 correspond to fertility postponement, with rates decreasing before age 25 and increasing at higher ages.

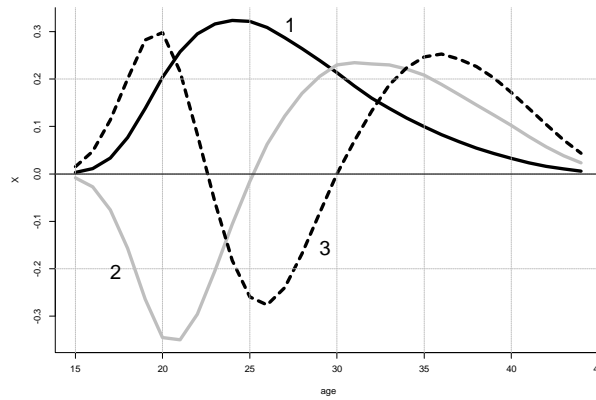


Figure 2. First three principal components  $\mathbf{X}$ , from the singular value decomposition of historical cohort schedules. These three components account for > 95% of deviations of historical rates from their age-specific means.

Any cohort schedule  $\theta_c$  can be decomposed into a part lying in the column space of  $\mathbf{X}$  and an orthogonal remainder:

$$(7) \quad \theta_c = \mathbf{X}\beta_c + \varepsilon_c = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\theta_c + \varepsilon_c$$

where the remainder vector is

$$(8) \quad \varepsilon_c = \left[ \mathbf{I}_A - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right] \theta_c = \mathbf{M}\theta_c$$

Because  $\mathbf{X}$  contains schedule components with large singular values, remainder vectors should typically be small. We can define “small” by constructing residual vectors for all complete cohort schedules in the historical array  $\mathbf{HIST}$ , and calculating their average outer product:

$$(9) \quad \bar{\mathbf{\Omega}} = \frac{1}{S} \sum_s \boldsymbol{\varepsilon}_s \boldsymbol{\varepsilon}_s'$$

These historical data allow us to establish a scalar penalty for the “badness” of each cohort schedule’s shape

$$(10) \quad \begin{aligned} \pi_c &= \boldsymbol{\varepsilon}_c' \bar{\mathbf{\Omega}}^+ \boldsymbol{\varepsilon}_c \\ &= \boldsymbol{\theta}_c' [\mathbf{M} \bar{\mathbf{\Omega}}^+ \mathbf{M}] \boldsymbol{\theta}_c \\ &= \boldsymbol{\theta}' [\mathbf{G}_c' \mathbf{M} \bar{\mathbf{\Omega}}^+ \mathbf{M} \mathbf{G}_c] \boldsymbol{\theta} \\ &= \boldsymbol{\theta}' \mathbf{K}_c \boldsymbol{\theta} \end{aligned}$$

where the + superscript represents the Moore-Penrose pseudo-inverse (Penrose 1955), necessary because  $\text{rank}(\mathbf{M})=A-3=27$ . By construction, the empirical average of  $\pi_c$  across the cohort schedules in **HIST** equals 27.

In our forecasting problem we deal with surfaces, like that in Figure 1, for which we have precise rate estimates at every age for the first 10 cohorts 1956...1965. We therefore use the shape penalties for cohorts with at least some unknown rates, namely for the 30 cohorts born in 1966...1995.

An important feature of this cohort shape penalty  $\pi_c$  is that it is improper, in the sense that an uncountably infinite number of fertility schedules correspond to any given level of the penalty. To take the simplest example, the minimum penalty  $\pi_c=0$  occurs for *any* schedule that is an exact linear combination of  $\mathbf{X}$  columns, regardless of the specific weights on the columns. In other words, by applying this penalty we assume no *a priori* knowledge of the specific shapes or levels of cohort fertility schedules. We assume only that, in the cohort dimension, a surface  $\theta$  can be well-approximated by the same components that best approximate historical schedules. An

important benefit of this approach is that a rate surface could have cohort schedules with shapes and levels not seen in the historical data, without heavy penalties.<sup>4</sup>

Figure 3 illustrates the shape penalty. It contains the observed cohort schedule for USA 1942-born women from the HFD (dark solid line), the projection of that schedule onto the column space of SVD components  $\mathbf{X}$  (solid squares), and the residuals  $\varepsilon$  that cannot be explained via the  $\mathbf{X}$  components (thick grey line along horizontal axis). After calculating the overall covariance matrix of shape residuals as in equation (9), the calculated penalty for this observed schedule is  $\pi_c=27.07$ . This penalty is very close to the empirical average of 27, so that the size and pattern of the USA 1942 residuals are in some sense typical of historical data. The most interesting feature in Figure 3 are the three dashed lines, which represent other hypothetical schedules with identical penalties  $\pi_c=27.07$ . By design, our improper cohort shape penalty cannot distinguish any of these four very different age patterns as more or less likely than any other.

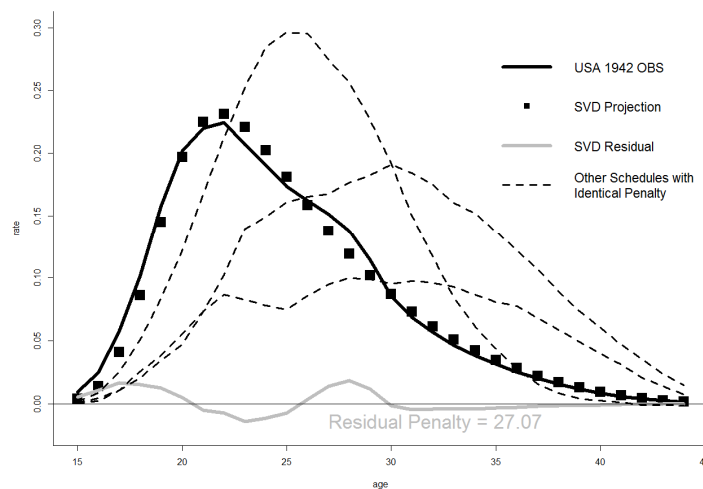


Figure 3. Observed cohort fertility schedule for US women born in 1942 (solid dark line), and best approximation of that schedule using the 3 SVD components in Figure 2 (dark squares). The approximation residuals, represented by a thick grey line, generate a shape penalty of 27.07. Dashed lines represent hypothetical schedules with identical values for shape penalty.

<sup>4</sup> Because there are no priors on the component weights, a three-component approach is flexible enough to allow many shapes that are not well represented in the historical data. Perhaps most importantly, our measure does not heavily penalize the bimodal age patterns that may be emerging in English-speaking countries (Sullivan 2005).

*Time Series Penalties*

Our second set of penalties concerns change and stability in the time series of rates at each age. Current demographic forecasting models use two main methods for extrapolating observed fertility rates into the future, which we will call the *freeze-rate* and *freeze-slope* approaches. The freeze-rate method assumes that the most likely future value for the fertility rate at age  $a$  is simply the last observed rate at that age. The freeze-slope method assumes that trends, measured as fitted slopes over some recent period, will continue into the future. In terms of residuals, these two methods suggest that

$$\text{Freeze-Rate: } \theta_{a,c+1} - \theta_{a,c} \approx 0$$

$$\text{Freeze-Slope: } \theta_{a,c+1} - \left[ \theta_{a,c} + \hat{\beta}(\theta_{a,c}, \dots, \theta_{a,c-n}) \right] \approx 0$$

where  $\hat{\beta}()$  is a slope estimator based on the recent history of rates at age  $a$ . As with the shape residuals in the previous subsection, one can construct penalties for a given rate surface, based on standardized residuals from rules of this type. The larger these time series residuals are, the less plausible is the rate surface  $\theta$  *a priori*.

There is wisdom in both of the standard forecasting approaches: age-specific rates do trend steadily upward or downward over periods of five or ten years (favoring *freeze-slope*), but biological constraints and the impossibility of negative rates also mean that such trends cannot continue indefinitely (favoring *freeze-rate*). In a Bayesian framework a researcher does not have to make an explicit choice between these competing models. Because the models are not mutually exclusive (a sequence of rates can be both constant and smooth) we can incorporate both probabilistically, and then calibrate the errors appropriately.

At each age, we define a vector of 30 freeze-rate residuals for cohorts 1966...1995:

$$(11) \quad u_a = \begin{bmatrix} \theta_{a,1966} - \theta_{a,1965} \\ \vdots \\ \theta_{a,1995} - \theta_{a,1994} \end{bmatrix} = \begin{bmatrix} 0 & \cdots & -1 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & -1 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \theta_a = \mathbf{W}_R \theta_a = \mathbf{W}_R \mathbf{H}_a \theta$$

and a similar vector of 30 freeze-slope residuals

$$(12) \quad v_a = \begin{bmatrix} \theta_{a,1966} - \left[ \theta_{a,1965} + \hat{\beta}(\theta_{a,1965}, \dots, \theta_{a,1961}) \right] \\ \vdots \\ \theta_{a,1995} - \left[ \theta_{a,1994} + \hat{\beta}(\theta_{a,1994}, \dots, \theta_{a,1990}) \right] \end{bmatrix} = \mathbf{W}_S \theta_a = \mathbf{W}_S \mathbf{H}_a \theta$$

For the freeze-slope case we estimate a local regression slope from the model  $(\theta_{a,c} - \theta_{a,c-n}) = \beta n$ ,

by least squares fitting over a five-year period using  $n=0 \dots 4$ . This produces slope estimates

$$\hat{\beta} = \frac{10}{30} \theta_{a,c} - \frac{1}{30} \theta_{a,c-1} - \frac{2}{30} \theta_{a,c-2} - \frac{3}{30} \theta_{a,c-3} - \frac{4}{30} \theta_{a,c-4}$$

and freeze-slope residuals that are weighted

sums of  $\theta$ s, such as  $v_{a,1966} = (\theta_{a,1966} - \frac{40}{30} \theta_{a,1965} + \frac{1}{30} \theta_{a,1964} + \frac{2}{30} \theta_{a,1963} + \frac{3}{30} \theta_{a,1962} + \frac{4}{30} \theta_{a,1961})$  for the

1966-born cohort. These more complex  $\beta$ -weights appear in the appropriate cells of the  $\mathbf{W}_S$  matrix.

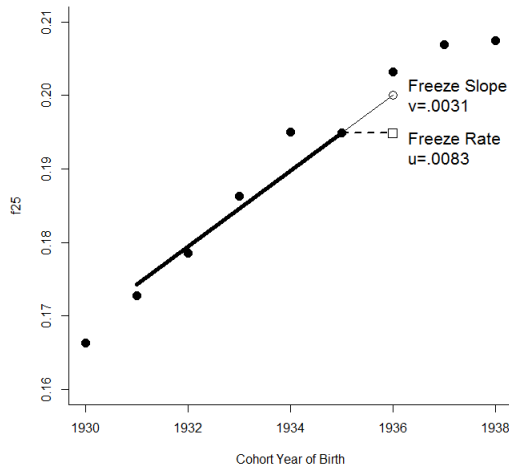


Figure 4. Freeze-rate residuals (u) and Freeze-slope residuals (v) at age 25, for 1936-born women in the Netherlands.

Figure 4 shows a simple example: freeze-rate and freeze-slope residuals at age 25, for women in the Netherlands who were born in 1936. In this particular example both residuals are small, as they typically are in the historical data, given the smoothness and short-term predictability of rates at any given age across adjacent cohorts. Both residuals are positive because the two models underpredict the observed rate. In this specific case the freeze-slope residual is smaller in magnitude, but one can see that for other cohorts (e.g. 1938) the freeze-rate residual is smaller. Figure 4 illustrates residual calculation at age 25 for a single cohort; each  $u_a$  and  $v_a$  vector contain 30 such residuals, for cohorts 1966-1995.

We calibrated the time series penalties by estimating all freeze-rate and freeze-slope residuals at each age in the historical data. The mean residuals of both methods are near zero at all ages, so that the average squared residuals for each (age, method) combination serve as estimates of residual variance – call these empirical estimates  $s_{Ra}^2$  and  $s_{Sa}^2$ . As we did for shape residuals, we standardize before constructing penalties. The freeze-rate penalty at age  $a$  therefore equals

$$\begin{aligned}
 \pi_{Ra} &= s_{Ra}^{-2} u_a' u_a \\
 (13) \quad &= \theta' \left[ s_{Ra}^{-2} \mathbf{H}'_a \mathbf{W}'_R \mathbf{W}_R \mathbf{H}_a \right] \theta \\
 &= \theta' \mathbf{K}_{Ra} \theta
 \end{aligned}$$

and the analogous freeze-slope penalty is

$$\begin{aligned}
 \pi_{Sa} &= s_{Sa}^{-2} v_a' v_a \\
 (14) \quad &= \theta' \left[ s_{Sa}^{-2} \mathbf{H}'_a \mathbf{W}'_S \mathbf{W}_S \mathbf{H}_a \right] \theta \\
 &= \theta' \mathbf{K}_{Sa} \theta
 \end{aligned}$$

It is useful to contrast these penalties with common Bayesian priors for smoothness over age and time (e.g., Breslow and Clayton 1993; Berzuini and Clayton 1994; Ogata et al. 2000; Bray 2002; Schmid and Held 2004; Girosi and King 2008). In general, smoothness priors assign



high probabilities to series with slowly changing slopes. The most common model in the literature (called *RW2* by Schmid and Held 2004) assumes that second differences in a series follow a random walk with small, independent perturbations. *RW2* is therefore equivalent to a freeze-slope model in which each time series value is predicted with error from the two preceding values (Breslow and Clayton 1993:17-18, Berzuini and Clayton 1994:Fig. 3). When examining the historical data, we found that *RW2* models that were appropriately smooth (i.e., with expected one-year-ahead prediction residuals similar to our historical data) were also too volatile (expected five- or ten-year-ahead differences were much larger than the corresponding historical averages). As a consequence, we adopted the variant described above: we use five, rather than two, years of earlier data to estimate the expected slope, and we also add freeze-rate penalties in order to identify less-volatile series as more plausible *a priori*.

#### *Weighting Multiple Shape and Time Penalties in the Prior Distribution*

The complex matrix notation in the previous subsection tends to obscure a relatively simple structure, so it is useful to pause and remember that there are three basic categories of *a priori* information, and that each penalty term has been standardized using empirical variance information from pre-1950 cohorts. By using this information to construct a prior distribution for  $\theta$ , we implicitly assume that general features of past rate surfaces (measured in terms of the frequencies of different values for shape and time series penalties) will persist into the future in the countries for which we will forecast age-specific rates and CFRs.

Table 1 summarizes the logic and the notation that we have developed so far regarding penalties over a 30x40 surface of fertility rates for ages 15...44 and cohorts 1956...1995.

<b>Table 1. Summary of Sub-Penalties for 30x40 Rate Surfaces</b>			
	<b>Schedule Shapes</b>	<b>Time-Series (Freeze Rate)</b>	<b>Time-Series (Freeze Slope)</b>
# of Penalties	30	30	30
Penalty Terms	$\pi_{1966} \dots \pi_{1995}$	$\pi_{R,15} \dots \pi_{R,44}$	$\pi_{S,15} \dots \pi_{S,44}$
Residuals	$\boldsymbol{\epsilon}_c = \mathbf{M} \boldsymbol{\theta}_c$	$\mathbf{u}_a = \mathbf{W}_R \boldsymbol{\theta}_a$	$\mathbf{v}_a = \mathbf{W}_S \boldsymbol{\theta}_a$
Penalty Matrices	$\mathbf{K}_{1966} \dots \mathbf{K}_{1995}$	$\mathbf{K}_{R,15} \dots \mathbf{K}_{R,44}$	$\mathbf{K}_{S,15} \dots \mathbf{K}_{S,44}$
<i>A Priori</i> Assumption	Schedules well approximated by SVD basis functions $\mathbf{X}$	Next cohort's rate at age $a$ well predicted by current rate	Next cohort's rate at age $a$ well predicted by recent trend
Calibration Information from Historical Data	Projection errors from $\mathbf{X}$	One-ahead freeze-rate prediction errors	One-ahead freeze-slope prediction errors
# of elements in each residual	30	30	30
Expected value of each penalty (=rank of $\mathbf{M}$ or $\mathbf{W}$ )	27	30	30

Our prior distribution additively combines all 90 of the penalty terms in its log likelihood, using a set of weights to adjust the contribution of each penalty  $j=1 \dots 90$ .

$$\begin{aligned}
 \ln f(\boldsymbol{\theta}, \mathbf{w}) &= \text{const} - \frac{1}{2} \sum_j w_j \pi_j \\
 (15) \qquad \qquad \qquad &= \text{const} - \frac{1}{2} \boldsymbol{\theta}' \left( \sum_j w_j \mathbf{K}_j \right) \boldsymbol{\theta} \\
 &= \text{const} - \frac{1}{2} \boldsymbol{\theta}' \mathbf{K} \boldsymbol{\theta}
 \end{aligned}$$

Non-unit weights are necessary because the residuals on which we base the penalties are not mutually independent – merely as an example, if cohort shape residuals for a surface are all very small, then large time series residuals may be less likely. We demonstrate in the Appendix that for a weighted prior distribution, the expected value of the  $j$ th penalty given  $\{w_j\}$  is

$$(16) \qquad \qquad \qquad E^*(\pi_j | \mathbf{w}) = \text{trace}(\mathbf{K}_j \mathbf{K}^+)$$

where  $E^*$  is a special operator for the expectation when  $\boldsymbol{\theta}$  is restricted to the column space of  $\mathbf{K}$  (see Girosi and King 2008), and  $\mathbf{K}^+$  is the generalized Moore-Penrose inverse of the weighted sum  $\mathbf{K}$ .

Our strategy for selecting weights to match historical data is as follows. By construction, we know the empirical average for each penalty term in the historical data: 27 for cohort shape penalties and 30 for time series penalties. Our objective is to find weights  $w_1 \dots w_{90}$  such that

$$(17) \quad E_*(\pi_j | w) = \text{trace}(\mathbf{K}_j \mathbf{K}^+) = \text{target}_j \quad \text{for } j = 1 \dots 90$$

where  $\text{target}_j$  represents the historical average for the penalty.

In practice, the following elementary search procedure converged quickly to good solutions:

0. Initialize all weights at unity:  $w_1 = w_2 = \dots = w_{90} = 1$
1. Calculate  $\mathbf{K} = \sum w_j \mathbf{K}_j$ , and its generalized inverse  $\mathbf{K}^+$
2. Calculate  $E_*(\pi_j | w) = \text{trace}(\mathbf{K}_j \mathbf{K}^+)$  for all  $j = 1 \dots 90$
3. Update weights as  $w_j^{new} = w_j \cdot \frac{E_*(\pi_j | w)}{\text{target}_j}$   $j = 1 \dots 90$
4. Stop if converged; otherwise return to step 1

Table 2 summarizes the results of the joint weighting procedure, showing the range of weights and expected values of the penalties before and after 30 iterations. These weights produce a  $\mathbf{K}$  matrix for a prior distribution for which the *a priori* expected values of each penalty match the historical average very closely.

**Table 2. Iterative Penalty Weighting**

	Schedule Shapes <b>27</b>	Time-Series (Freeze Rate) <b>30</b>	Time-Series (Freeze Slope) <b>30</b>
<b>Target value <math>E^*(\pi w)</math></b>			
<b>Range of w</b>			
Before Iteration 1	1.000 – 1.000	1.000 – 1.000	1.000 – 1.000
After Iteration 30	0.644 – 0.886	0.069 – 0.535	0.338 – 0.600
<b>Range of <math>E^*(\pi w)</math></b>			
Before Iteration 1	5.611 – 15.304	5.611 – 15.304	10.238 – 14.913
After Iteration 30	<b>27.000 – 27.000</b>	<b>29.997 – 30.000</b>	<b>30.000 – 30.001</b>

Joint calibration produces the model that we use for forecasting, via the conditional posterior normal described in Equation (6):

$$(18) \quad \theta | y \sim N \left\{ \begin{array}{l} \mu_{post} = [\mathbf{V}'\Psi^{-1}\mathbf{V} + \mathbf{K}]^{-1} [\mathbf{V}'\Psi^{-1}y] , \\ \Sigma_{post} = [\mathbf{V}'\Psi^{-1}\mathbf{V} + \mathbf{K}]^{-1} \end{array} \right\}$$

where  $\mathbf{K}$  is the weighted sum from Equation (15) using the jointly calibrated weights.

The *maximum a posteriori* (MAP) estimator of true rates  $\theta$  is the  $CA \times 1$  mean vector  $\mu_{post}$ , and the  $CA \times CA$  matrix  $\Sigma_{post}$  quantifies posterior uncertainty. The  $n$  elements of  $\theta$  that occur before the forecast date will have very low posterior variances, because we have precise HFD estimates for those rates from national statistical agencies. The remaining  $CA-n$  elements comprise the forecast.

### *Improper Priors and CFR Forecasts*

We emphasize two important points about the historically-calibrated prior distribution. First, it captures features of fertility surfaces that are remarkably robust to changes in the historical data from which it is constructed. In constructing and calibrating priors, we experimented with many subsets of the HFD, using different periods and different subsets of countries. In all cases the principal components in Figure 2, and the empirical error patterns in time series residuals, were quite similar. This consistency of results increases our confidence that the specific prior distribution that we have developed from the historical data truly captures robust qualitative patterns of fertility surfaces.

Second, the prior is completely uninformative about absolute fertility levels. The improper shape priors introduce information only about the relative levels of cohort fertility at different ages. Similarly, improper time series priors introduce information only about short-term

smoothness and stability of rates. For all but the youngest women (for whom we have little or no fertility history at the time of the forecast), CFR forecasts come mainly from the data, not from *a priori* assumptions.

### Coverage Validation & Comparisons with Alternative Models

As in any forecasting problem, it is important to understand the degree to which our model adequately estimates forecast uncertainty. In order to test coverage performance, we withheld post-1985 period data and simulated the forecasts that would have been produced in calendar year 1985.

An ideal simulation would move the entire forecast algorithm back in time by about 25 years: it would redefine the historical part of the HFD to include cohorts born 1875-1924, reconstruct and recalibrate a weighted penalty matrix  $\mathbf{K}$  from those data, and so on. Unfortunately, we lack sufficient historical data for this ideal procedure: only Sweden has complete data for any cohorts born before 1906, and for many countries in our set data collection began late enough that the earliest complete cohort was born after 1925.

Because of these data limitations, for the 1985 simulation we use the  $\mathbf{K}$  matrix derived for the 2010 forecasts in the previous section. This implies that, unlike in our actual forecasts, in the simulations there is some overlap between the data used to develop the prior and the own-country data used in the likelihood. The practical consequences of this overlap are slight: as mentioned previously, the improper prior is remarkably insensitive to the choice of training data.

Using the prior distribution based on  $\mathbf{K}$ , we calculated the posterior mean and covariance of  $\theta$  for each country over a grid including the 1931-1970 cohorts, based on data that would have been available in 1985. We then compared the posterior means and credibility intervals from this simulated forecast to the known fertility of those same cohorts over the next 25 years. Table 3

summarizes results and coverage of CFR forecasts for all of the countries for which we could produce 1985 forecasts, disaggregated by the age of the cohort at the 1985 forecast horizon.<sup>5</sup>

Table 3 shows imperfect, but reasonable, coverage for posterior credibility intervals. In particular, the 90% intervals perform well: they contain 92% of true future CFRs, with fairly good results by age at forecast. In contrast, 50% intervals are generally too wide for cohorts near the end of reproductive life at the forecast, and too narrow for cohorts who are 25 or 30 on the forecast date. The table also shows that most CFR forecast errors are small for cohorts that have completed fertility through age 25. In 1985, there would have been a small but definite bias in long-range forecasts of future fertility: for most of the cohorts then in their 20s, CFRs forecasts would have been slight underestimates. A Bayesian forecast with our relatively weak priors would have slightly over-extrapolated the downward trends in fertility across the cohorts born in the 1930s, because their continuation did not imply unusual shapes for cohort schedules. This problem occurred in the worst of the simulated 1985 forecasts, where the 90% posterior probability interval for the CFR of Portuguese women born in 1965 (and thus age 20 on the forecast date) was [1.11, 1.71], compared to a realized CFR of 1.83.

---

<sup>5</sup> Simulated forecast plots for all countries can be viewed at [WEBSITE://sim1985.html](http://WEBSITE://sim1985.html)

TABLE 3

Simulated 1985 Forecasts compared to observed\* CFR, by age of cohort in 1985.  
 Error = (Posterior MAP CFR – Observed CFR)

Age at Forecast	Mean Error	Mean Abs Error	% of Observations in Posterior Probability Intervals			
			5 to 95%ile	5 to 25%ile	25 to 75%ile	75 to 95%ile
20	-.09	.15	85	8	54	23
25	-.03	.09	81	13	31	38
30	+.01	.03	94	38	38	19
35	+.00	.01	100	19	75	6
40	+.00	.00	100	12	88	0
ALL	-.02	.05	92	18	57	17
Target	0	0	90	20	50	20

\*For coverage evaluation we treat CFR data from complete cohorts as observed constants. In fact they are very precise estimates from large national samples. The CFR<sub>1960</sub> column in Table 5 shows typical standard errors. Results are aggregated over 1985 simulations for the 15 countries in which a 1985 forecast was possible: Austria, Bulgaria, Canada, Czech Republic, Denmark, England & Wales, Finland, France, Hungary, Netherlands, Portugal, Slovakia, Sweden, Switzerland, USA.

It is also useful to compare alternative models, in order to evaluate the degree to which using both time and shape priors improves forecast coverage. Table 4 repeats the fourth column of Table 3, which reports the proportion of post-forecast observations falling in the 90% posterior interval for the full Bayesian model. The table also includes coverage calculations for two alternative forecast procedures. The first alternative model is a Bayesian model with time series priors only [i.e.,  $w_j=0$  for all cohort shape penalties in Equation (15)].<sup>6</sup> The second alternative is a forecast produced by fitting independent ARIMA(1,1,0) models to the time series of available rates at each age.

Coverage in a Bayesian model with only time series priors (Alternative 1) is notably worse than in the full model that includes shape penalties for cohort schedules. There is clearly considerable value added from the inclusion of shape priors that prioritize time trends that lead to more plausible shapes in the cohort dimension. A simple ARIMA model (Alternative 2) produces CFR confidence intervals that are too narrow for women who are younger than 35 on

<sup>6</sup> We omit a model with only a shape prior from the list of alternatives. Li and Wu (2003) noted that such models become unstable for women under 30 on the forecast date, with negative predicted rates at some ages and very large differences in the fitted schedules across adjacent cohorts. After confirming their observations in the historical data, we opted to exclude a shape-only model as a serious alternative.

the forecast date. This may occur because positive covariances in fertility rates at similar ages make the sum of age-specific rates (i.e., CFR) more variable than one would expect under an assumption of independent time series by age. The broad comparative coverage information in Table 4 does not highlight another problem with the independent ARIMA model – namely, a systematic negative bias in CFR forecast bands, with 84% of realized CFRs for the 1941-1966 cohorts falling above the posterior medians predicted by the ARIMA model, and 53% falling above the posterior 75<sup>th</sup> percentile.

Table 4. Coverage of three alternative simulated 1985 forecasts

Age at Forecast	Percent of CFR Observations* in 90% probability interval		
	Forecast Model	Alternative Model 1	Alternative Model 2
	Bayes (Shape+Time)	Bayes (Time Only)	ARIMA(1,1,0)
20	85	62	54
25	81	50	69
30	94	56	75
35	100	94	81
40	100	100	100
ALL	92	73	77
Target	90	90	90

\*See notes for Table 3.

We conclude from this simulation exercise that the Bayesian model with shape and time parameters is likely to perform well in terms of forecast coverage. It has good coverage properties across the 15 country forecasts that we were able to produce for 1985, and it clearly outperforms the tested alternatives.

### 2010 Fertility Forecasts from Contemporary Data

Using our model with the rate estimates available in 2010 for each country produces a joint posterior distribution for each fertility surface  $\theta$ , via Equation (17). With 24 countries, 30 ages, and hundreds of cohorts, this produces a very large set of output that we can only briefly



summarize in this article. We focus here on a few cases of interest. Readers can find a larger volume of summary graphics for the entire data set on our project website.<sup>7</sup>

One important feature of our model is the way in which priors for the shape of cohort schedules constrain the projected time series of age-specific rates. Linear time series extrapolations that would produce implausibly-shaped cohort schedules have low prior probability, so effectively the forecast must compromise between observed rate levels and trends in the period just before the forecast, and regularly-shaped fertility schedules for cohorts. As a result, projected trends in age-specific rates can sometime deviate substantially from freeze-rate or freeze-slope extrapolations.

Figure 5 shows an example, for Czech women at ages 25, 30, and 35. The solid points in the three time series correspond to horizontal slices across the surface in Figure 1, and illustrate dramatic changes in fertility timing. Prior information suggests that a combination of continued decline at age 25 with continued increases at ages 30 and 35 is very unlikely, however, because those changes would imply highly implausible shapes for the schedules of still-incomplete cohorts. In fact, the *maximum a posteriori* compromise between cohort shapes and trends for future Czech rates predicts fairly stable rates at ages 25, a reversal of recent increases at age 30, and less-than-linear increase in the rate at age 35.

---

<sup>7</sup> We have provided an offline copy of part of the website for reviewers, in the file *CohortFertility.zip*. To see the full set of results, unzip the file and open the *index.html* file in any browser. If the paper is published, we will make the full project web site, including all data and *R* code, available to readers on the web.

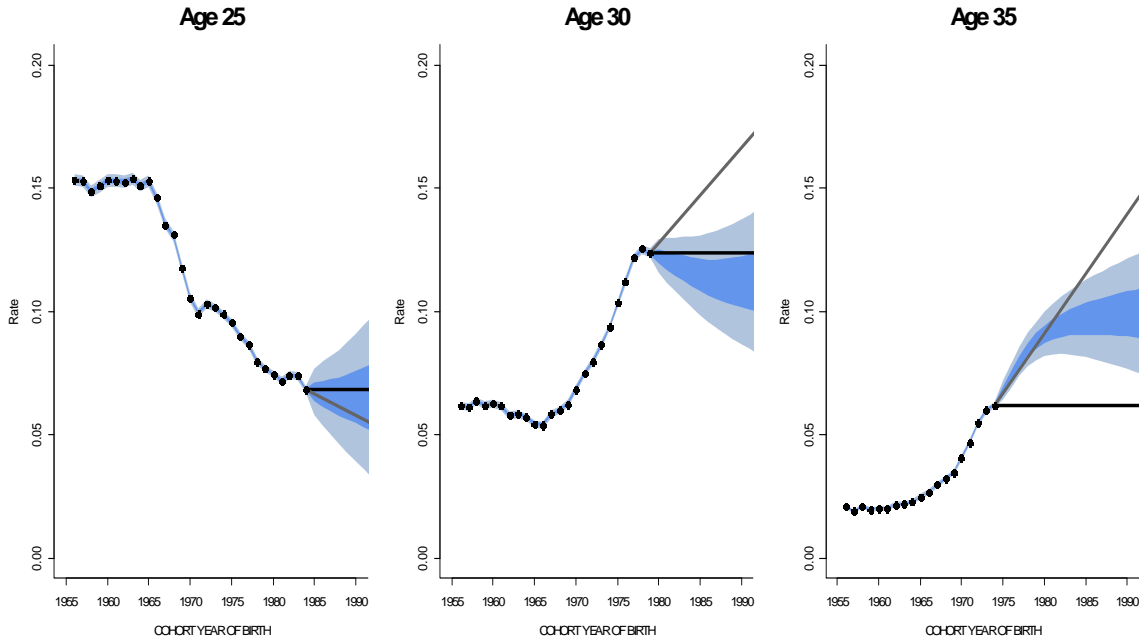


Figure 5. Czech fertility forecasts at ages 25, 30, and 35. Light and dark bands are 90% and 50% posterior probability intervals, respectively. Dots are HFD estimates, dark horizontal lines are freeze-rate forecasts using the last observed age-specific rate, grey solid lines are freeze-slope forecasts based on last 5 observed rates.

If extrapolated linearly, Czech fertility for the 1995 cohort at age 30 would reach levels near the 90<sup>th</sup> percentile of all rates ever observed at that age, while rates at age 25 for that cohort would be below the 1<sup>st</sup> percentile.<sup>8</sup> Although our improper priors do not penalize a high level at age 30 or a low level at age 25 *per se*, they do say that combination is very unlikely, especially together with historically moderate rates at age 35. In short, *a priori* knowledge about relative fertility at different ages within cohorts implies, in this case, strong constraints on time patterns of age-specific rates across cohorts.

We have a special interest in completed cohort fertility. For cohort  $c$  the posterior distribution of completed fertility is

$$(19) \quad CFR_c = 1' \theta_c = 1' G_c \theta \sim N(1' G_c \mu_{post}, 1' G_c \Sigma_{post} G_c' 1)$$

<sup>8</sup>Linear extrapolation would eventually predict negative rates for any age group with a negative trend. Negative rates are also possible in our model, but in practice they turned out to be rare over our forecast period: of 42000 estimated rates for cohorts born 1956-1995, only 7 had negative posterior means and only 230 had 90% probability intervals containing zero.

from which one can calculate the MAP estimator and posterior probability intervals. More plausible patterns in cohort schedules and time series should translate into more plausible trends for completed cohort fertility.

Figure 6 illustrates four of the @35 CFR forecast series – for the USA, Netherlands, Czech Republic, and Singapore. The very narrow posterior probability intervals show that it is easy to forecast precisely the completed fertility for women who are already 30 and older, but much harder for younger women. Probability intervals for cohort CFR are extremely narrow for the cohorts with nearly complete fertility histories at the forecast date: biology makes it certain that cohorts of women in their late 30s and early 40s are already very close to their average completed family sizes.

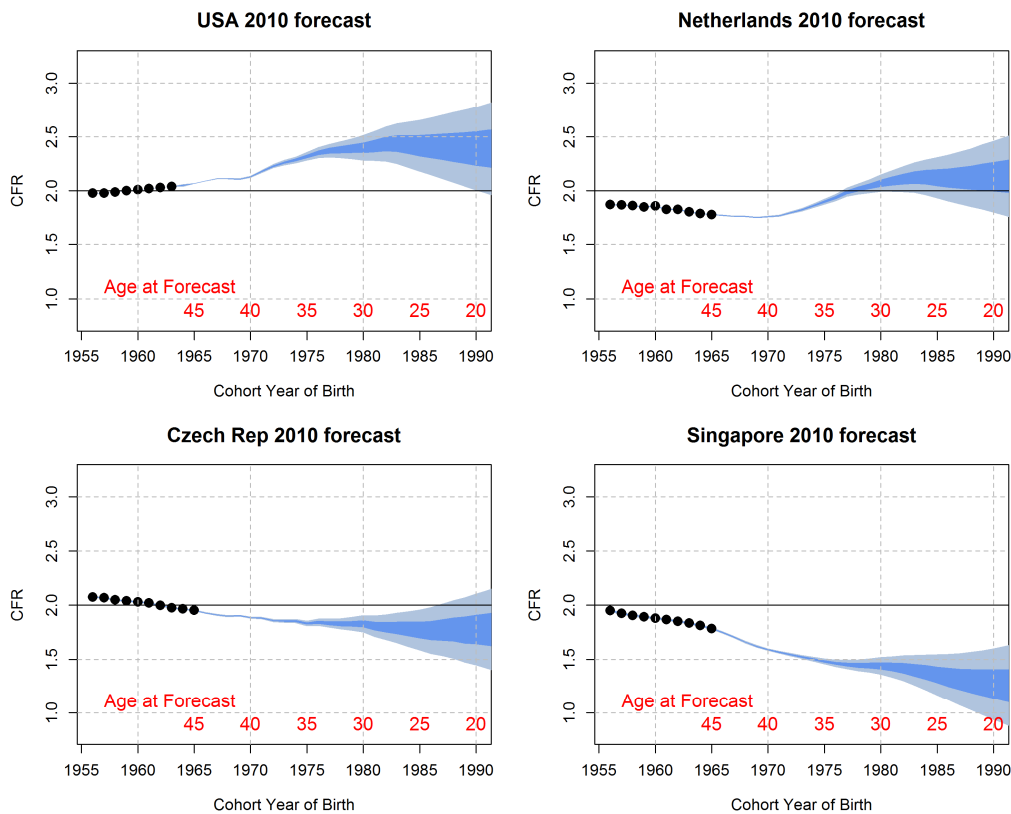


Figure 6. Posterior distributions of completed fertility -- USA, Netherlands, Czech Republic, and Singapore – conditional on rate estimates available in 2010. Light and dark bands are 90% and 50% posterior probability intervals, respectively. Dots are posterior estimates for cohorts with complete fertility histories through age 44.

Posterior uncertainty about the completed fertility of the later-born cohorts is small enough, however, to allow some important qualitative predictions about the likely fertility levels of women born in the 1980s in our sample of countries. A Bayesian approach allows us to make probabilistic statements about our main research questions: After steady declines in many countries, is cohort fertility likely to rebound or increase? The forecasts in Figure 6 provide visual answers: future increases are almost certain in the US (where completed CFRs are already rising), probable in the Netherlands, improbable in the Czech Republic (although the decline appears likely to stop), and almost certain not to occur in Singapore.

Table 5 presents a more systematic evaluation, showing for each country<sup>9</sup> the posterior mean of  $CFR_{1960}$ , and posterior means of the forecast differences between CFR for pairs of cohorts born 10 years apart. Values in the last three columns are positive if the later-born cohort is forecast to have higher completed fertility. Shaded cells have greater than 90% posterior probability of being positive (dark shading) or negative (light shading).

There are some unique country trajectories in Table 5. Most notably, Denmark and the US are the only countries in which women born in 1970 are likely to have more children on average than women born in 1960. The opposite holds in every other country in our data set. Note that these 1960-1970 changes are virtually certain, because women born in 1970 were already 40 in 2010, so that their final fertility levels can be forecast very precisely.

The most notable feature of Table 5 is the high probability of positive CFR change between the 1970 and 1980 cohorts in many countries. Our forecasts suggest that it is highly likely that in the near future, as women born in the 1970s reach their 45<sup>th</sup> birthdays, most countries will observe slight rebounds in completed fertility. Very few are likely to see continued

---

<sup>9</sup> In this table we report results for Germany as a whole only, and we omit Slovenia and Luxembourg because missing data for the 1956-1965 cohorts at young ages made a 2010 forecast impossible with our procedure.

decreases. However, taken as a whole the forecasts suggest that the decline in family sizes will probably stop, or even reverse itself, in many of the world's rich countries.

The general pattern does not apply everywhere, of course. CFR forecasts show continued decreases across cohorts born in the 1970s in several European countries (Portugal, Hungary, and Slovakia), in Asia (Singapore and Korea), and in Brazil (which is a special case because CFR is still falling due to rapid economic progress and a classical transition from high to near-replacement fertility levels).

Forecast uncertainty is much higher across the cohorts born in the 1980s, on whom we have shorter histories with little data on whether or not women are postponing childbearing. Although none of the 1970-1980 changes are significantly different from zero using the 90% probability threshold in Table 5, there are several countries (Sweden, Bulgaria, Russia) in which there is a fairly high posterior probability of sustained increase in average family sizes as women born in the 1980s 'cross the finish line' on their 45<sup>th</sup> birthdays. Similarly, continued decreases are fairly likely over the 1980s cohorts for Singapore, Portugal, Korea, Hungary, Estonia, Romania, and Brazil.

Table 5. Posterior means of CFR for women born in 1960, and for CFR changes between cohorts (standard deviations of changes in parentheses). Shaded cells have >90% probability of being negative (light shading) or positive (dark shading). Countries are sorted in ascending order of CFR<sub>1960</sub>.

Country	CFR <sub>1960</sub>	Expected CFR increase between birth cohorts		
		1960-1970	1970-1980	1980-1990
Germany	1.66 (.002)	-.15 (.00)	+.12 (.04)	-.07 (.17)
Italy	1.69 (.002)	-.21 (.00)	+.01 (.05)	+.02 (.18)
Austria	1.70 (.005)	-.07 (.01)	-.02 (.06)	-.08 (.20)
Switzerland	1.77 (.006)	-.12 (.01)	+.02 (.05)	-.04 (.18)
Canada	1.83 (.003)	-.03 (.01)	+.15 (.07)	+.01 (.21)
Japan	1.84 (.002)	-.36 (.00)	+.06 (.05)	-.01 (.18)
Russia	1.85 (.001)	-.24 (.00)	+.09 (.05)	+.08 (.18)
Netherlands	1.86 (.004)	-.10 (.01)	+.31 (.05)	+.06 (.18)
Belgium	1.87 (.005)	-.05 (.01)	+.23 (.05)	-.02 (.18)
Scotland	1.87 (.007)	-.11 (.01)	+.17 (.05)	+.04 (.18)
Denmark	1.88 (.007)	+.10 (.01)	+.18 (.04)	-.03 (.17)
Singapore	1.88 (.008)	-.29 (.01)	-.15 (.05)	-.17 (.18)
Lithuania	1.91 (.008)	-.16 (.01)	+.20 (.05)	.00 (.18)
Portugal	1.92 (.005)	-.24 (.01)	-.19 (.05)	-.11 (.18)
Bulgaria	1.95 (.006)	-.27 (.01)	+.17 (.05)	+.12 (.18)
Finland	1.96 (.007)	-.07 (.01)	+.12 (.05)	+.05 (.18)
Greece	1.96 (.005)	-.34 (.01)	-.05 (.04)	-.04 (.17)
England & Wales	1.97 (.002)	-.06 (.00)	+.21 (.05)	+.05 (.18)
USA	2.01 (.001)	+.12 (.01)	+.27 (.07)	-.01 (.21)
Hungary	2.02 (.005)	-.16 (.01)	-.27 (.05)	-.14 (.18)
Czech Republic	2.03 (.006)	-.14 (.01)	-.06 (.05)	-.05 (.18)
Sweden	2.05 (.006)	-.06 (.01)	+.15 (.04)	+.13 (.17)
Estonia	2.06 (.013)	-.18 (.02)	.00 (.05)	-.18 (.18)
Korea	2.08 (.002)	-.33 (.00)	-.07 (.04)	-.07 (.17)
France	2.11 (.002)	-.11 (.00)	+.17 (.05)	-.04 (.18)
Romania	2.16 (.004)	-.54 (.01)	+.05 (.05)	-.08 (.18)
Australia	2.17 (.004)	-.14 (.01)	+.07 (.05)	-.04 (.18)
Slovakia	2.17 (.007)	-.24 (.01)	-.21 (.05)	+.03 (.18)
New Zealand	2.37 (.009)	-.19 (.01)	+.13 (.04)	+.07 (.17)
Northern Ireland	2.42 (.014)	-.29 (.02)	+.23 (.05)	-.06 (.18)
Iceland	2.45 (.034)	-.13 (.04)	+.12 (.06)	-.04 (.18)
Brazil	2.86 (.002)	-.48 (.00)	-.13 (.04)	-.10 (.17)

As the widening probability intervals in Figure 6 and the increasing posterior standard deviations in Table 5 make clear, forecasts become far more speculative as we move to later-born cohorts on whom we currently have shorter fertility histories. A great advantage of Bayesian modeling over many other demographic forecasting methods is that we can quantify the uncertainty about these speculative results.

## **Discussion**

Over the last decade period fertility rates have risen in many developed countries, in part due to a decelerating shift of births to older maternal ages. The impact of these changes on cohort fertility is unclear, because the cohorts responsible for most births during this time of recent increase are still ten to twenty years from completing their childbearing. To know if cohorts that have postponed childbearing will ultimately have fewer children, one needs to forecast, but forecasting fertility is notoriously difficult (Booth 2006).

We have developed new Bayesian forecasting methods for completed cohort fertility, and applied them to the countries in the Human Fertility Database and to a number of additional countries. Past efforts to forecast completed cohort fertility have typically relied exclusively on time trends, or on parametric models for rate schedules. We combine these two approaches, borrowing strength from recent time trends and also from historical patterns in cohort age profiles. Our forecasts suggest that cohort fertility is likely to stabilize or even increase slightly in several countries. Among women born in the 1970s and early 1980s, current rate trends suggest historically plausible age patterns of fertility that would lead to slightly larger average family sizes for the women born later. The pattern appearing recently in several Scandinavian countries (Andersson et al. 2009) may be spreading elsewhere.

Our method not only provides forecasts, but also quantifies uncertainty. From the column labeled “1970-80” in Table 5, for example, we see that for women born in the 1970s, downward trends in cohort fertility appear to be slowing or reversing in many countries. The youngest of these women are not quite 30 at our forecast horizon, but posterior probabilities show that we know enough to predict with high confidence that their completed fertility will be greater than that of women born 10 years earlier. For women born in the 1980s, the oldest of whom were in their late 20s at the forecast horizon, forecasts become much more uncertain and the posterior distribution tells us so: from standard deviations one can see that probability intervals for  $CFR_{1990}-CFR_{1980}$  in the third column of Table 5 would be approximately three to four times as wide as those for the  $CFR_{1980}-CFR_{1970}$  difference in the second column.

Quantifying uncertainty also shows that one can predict the final fertility of older cohorts very precisely. This occurs not only because completing the fertility of older women involves a short forecast period, but also because their future fertility paths are very well known *a priori*. The very narrow probability intervals shown in Figure 6 for cohorts born as recently as 1980 indicate that demographers already know quite a lot about the average completed fertility of women who are today in their 30s.

Our methods are quite general. The same approach could be applied, with only minor modifications, to many demographic problems. These include forecasting childlessness and parity-specific fertility, and forecasts of first marriage and never-marrying. Beyond demography, the technique of SVD decomposition with penalized projection residuals is adaptable across many domains. The idea of using simultaneous penalties in overlapping dimensions is similarly applicable to many problems. Quadratic penalties and multivariate normality could be used in a variety of forecasting and other missing-data problems. The quadratic penalty approach is



computationally convenient, because it does not require extensive sampling from the posterior distribution. It is also relatively easy to explain to researchers who are unfamiliar with Bayesian vocabulary, because posterior means can also be explained as penalized least-squares solutions, as ridge regressions, or even as variants of smoothing splines (Wood 2000).

Probabilistic forecasting methods have become standard in mortality forecasting, since the introduction of Lee and Carter's (1992) approach combining singular value decomposition of rates by age and time with classical time series methods. Girosi and King (2008) and Soneji and King (2011) have recently used Bayesian methods to add additional structure to mortality forecasts, by using prior information on age patterns to make forecasts demographically more coherent. Bayesian models also allow forecasts from sparser and lower-quality data, such as that available for cause of death.

Although uncertainty is inherently greater for fertility forecasts, statistical agencies and demographers studying cohort trends have generally relied on deterministic projections and forecasts. Our hope is that developments in probabilistic fertility modeling, in combination with the public availability of high quality data sets like the Human Fertility Database, will increase understanding of fertility trends and help demographic researchers to express more precisely how much we do – and don't – know about the future.

## References

- L Alkema, AE Raftery, P Gerland, SJ Clark, F Pelletier, T Buettner, and GK Heilig, 2011. "Probabilistic projections of the total fertility rate for all countries". *Demography* 48(3):815-839.
- G Andersson, M Ronsén, LB Knudsen, T Lappegård, G Neyer, K Skrede, K Teschner, and A Vikat. 2009. "Cohort fertility patterns in the Nordic countries". *Demographic Research* 20(14):313-352.
- C Berzuini and D Clayton, 1994. "Bayesian analysis of survival on multiple time scales". *Statistics in Medicine* 13:823-838.
- J de Beer, 1985. "A time series model for cohort data". *Journal of the American Statistical Association* 80(391):525-530.
- DE Bloom, 1982. "What's happening to the age at first birth in the United States? A study of recent cohorts". *Demography* 19(3):351-370.
- J Bongaarts and G Feeney, 1998. "On the quantum and tempo of fertility". *Population and Development Review* 24(2):271-291.
- H Booth, 2006. "Demographic forecasting: 1980 to 2005 in review". *International Journal of Forecasting* 22:547-581.
- H Booth and L Tickle, 2008. "Mortality modeling and forecasting: a review of methods". ADSRI Working Paper No. 3. Australian National University.
- I Bray, 2002. "Application of Markov Chain Monte Carlo methods to projecting cancer incidence and mortality". *Journal of the Royal Statistical Society C* 51(2):151-164.
- NE Breslow and DG Clayton, 1993. "Approximate inference in generalized linear mixed models". *Journal of the American Statistical Association* 88:9-25.
- R Chen and SP Morgan, 1991. "Recent trends in the timing of first births in the United States". *Demography* 28(4):513-533.
- Y-H A Chen, 2010. "The future of completed cohort fertility in low fertility countries". Population Association of America Meeting. Dallas, TX.
- P-C R Cheng and ES Lin, 2010. "Completing incomplete cohort fertility schedules". *Demographic Research* 23/9:223-256.
- T Frejka and G Calot. 2001. "Cohort reproductive patterns in low-fertility countries". *Population and Development Review* 27:103-132.
- F Girosi and G King, 2008. *Demographic Forecasting*. Princeton University Press.

JR Goldstein, 2008. "A behavioral Gompertz model for cohort fertility schedules in low and moderate fertility populations". Population Association of America Meeting. New Orleans LA.

JR Goldstein, T Sobotka, and A Jasilioniene, 2009. "The End of Lowest-low Fertility?". *Population and Development Review* 35(4):663-699.

Human Fertility Database (HFD), 2011. Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria). Available at [www.humanfertility.org](http://www.humanfertility.org) (data downloaded on 2 Nov 2011).

RJ Hyndman and H Booth, 2008. "Stochastic population forecasts using functional data models for mortality, fertility and migration". *International Journal of Forecasting* 24(3):323-342.

H-P Kohler and D Philipov, 2001. "Variance effects in the Bongaarts-Feeney formula". *Demography* 38:1-16.

RD Lee, 1993. "Modeling and forecasting the time series of US fertility: age, distribution, range, and ultimate level". *International Journal of Forecasting* 9:187-202.

RD Lee and LR Carter, 1992. "Modeling and forecasting US mortality". *Journal of the American Statistical Association* 87(419):659-671

N Li and Z Wu, 2003. "Forecasting Cohort Incomplete Fertility: A Method and an Application". *Population Studies* 57(3): 303-320.

A Luci and O Thevenon, 2010. "Does economic development drive fertility rebound in OECD countries?". Population Association of America Annual Meeting. Dallas, TX.

H Lütkepohl, 2006. *New introduction to multiple time series analysis*. Springer.

M Myrskylä, JR Goldstein, and YA Chen, 2013. "New cohort fertility forecasts for the developed world: rises, falls, and reversals". *Population and Development Review* 39(1):31-56.

M Myrskylä, H-P Kohler, and FC Billari, 2009. "Advances in development reverse fertility declines". *Nature* 460(7256):741-743.

Y Ogata, K Katsura, N Keiding, C Holst and A Green, 2000. "Empirical Bayes age-period-cohort analysis of retrospective incidence data". *Scandinavian Journal of Statistics* 27(3):415-432.

R Penrose, 1955. "A generalized inverse for matrices". *Mathematical Proceedings of the Cambridge Philosophical Society* 51:406-413.

AE Renshaw and S Haberman, 2006. "A cohort-based extension to the Lee-Carter model for mortality reduction factors". *Insurance: Mathematics and Economics* 38(3):556-570

V Schmid and L Held, 2004. "Bayesian extrapolation of space-time trends in cancer registry data". *Biometrics* 60(4):1034-1042.

T Sobotka, 2008. "The rising importance of migrants for childbearing in Europe". *Demographic Research* 19(9):Overview Chapter 7.

S Soneji and G King, 2011. "The future of death in America". *Demographic Research* 25/1:1-38. doi:10.4054/DemRes.2011.25.1

R Sullivan, 2005. "The age pattern of first-birth rates among US women: the bimodal 1990s". *Demography* 42(2):259-273.

SN Wood, 2000. "Modelling and smoothing parameter estimation with multiple quadratic penalties". *Journal of the Royal Statistical Society B* 62(2):413-428.

PA Thompson, WR Bell, JF Long, and RB Miller, 1989. "Multivariate time series projections of parameterized age-specific fertility rates". *Journal of the American Statistical Association* 84(407):689-699.

E van Imhoff and N Keilman, 2000. "On the quantum and tempo of fertility: Comment". *Population and Development Review* 26(3):549-553.

Y Zeng and KC Land, 2002. "Adjusting period tempo changes with an extension of Ryder's basic translation equation". *Demography* 39:269-285.

## Appendix

As described in Girosi and King (2008), an improper prior of the form

$$\ln f(\boldsymbol{\theta}) = \text{const} - \frac{1}{2} \boldsymbol{\theta}' \mathbf{K} \boldsymbol{\theta}$$

can be understood through eigen-decomposition of the rank-deficient matrix  $\mathbf{K}$ . Specifically, write that decomposition as

$$\mathbf{K} = \mathbf{U} \mathbf{D} \mathbf{U}' = [\mathbf{U}_1 \quad \mathbf{U}_0] \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{U}_1' \\ \mathbf{U}_0' \end{bmatrix} = \mathbf{U}_1 \mathbf{D}_1 \mathbf{U}_1'$$

where  $\mathbf{U}$ 's columns are orthonormal eigenvectors,  $\text{rank}(\mathbf{K})=r$ ,  $\mathbf{U}_0$  contains eigenvectors corresponding to the zero eigenvalues,  $\mathbf{D}_1$  is an  $r \times r$  diagonal matrix of positive eigenvalues, and  $\mathbf{U}_1$  is a  $1200 \times r$  matrix containing the eigenvectors corresponding to the positive eigenvalues.

Adopting a new orthogonal coordinate system ( $\gamma$ ) based on the columns of  $\mathbf{U}$  yields

$$\boldsymbol{\theta} = [\mathbf{U}_0 \quad \mathbf{U}_1] \begin{bmatrix} \gamma_0 \\ \gamma_1 \end{bmatrix} = \mathbf{U}_0 \gamma_0 + \mathbf{U}_1 \gamma_1$$

where  $\gamma_0 = \mathbf{U}_0' \boldsymbol{\theta}$  and  $\gamma_1 = \mathbf{U}_1' \boldsymbol{\theta}$ . In terms of the  $\gamma$ -coordinates, a prior based on  $\mathbf{K}$  implies

$$\begin{aligned} \ln f(\boldsymbol{\gamma}) &= \text{const} - \frac{1}{2} (\gamma_0' \mathbf{U}_0' + \gamma_1' \mathbf{U}_1') (\mathbf{U}_1 \mathbf{D}_1 \mathbf{U}_1') (\mathbf{U}_0 \gamma_0 + \mathbf{U}_1 \gamma_1) \\ &= \text{const} - \frac{1}{2} \gamma_1' \mathbf{D}_1 \gamma_1 \end{aligned}$$

In other words, the improper prior based on  $\mathbf{K}$  tells us that  $\boldsymbol{\theta}$ 's  $\gamma_1$  coordinates have a normal distribution with mean zero and covariance matrix  $\mathbf{D}_1^{-1}$ , while the remaining  $(1200-r)$   $\gamma_0$  coordinates are completely unrestricted.

Under a weighted, combined prior with  $\mathbf{K} = \sum_j (w_j \mathbf{K}_j)$ , the  $j$ -th penalty is

$$\begin{aligned} \pi_j &= \boldsymbol{\theta}' \mathbf{K}_j \boldsymbol{\theta} \\ &= (\gamma_0' \mathbf{U}_0' + \gamma_1' \mathbf{U}_1') \mathbf{K}_j (\mathbf{U}_0 \gamma_0 + \mathbf{U}_1 \gamma_1) \\ &= \gamma_1' \mathbf{U}_1' \mathbf{K}_j \mathbf{U}_1 \gamma_1 + \text{terms involving } \gamma_0 \end{aligned}$$

If we define a special expectation operator  $E^*$  that always conditions on  $\gamma_0=0$ , then it is possible

to calculate

$$\begin{aligned}
 E^*(\boldsymbol{\pi}_j | w) &= E^*(\boldsymbol{\gamma}'_1 \mathbf{U}'_1 \mathbf{K}_j \mathbf{U}_1 \boldsymbol{\gamma}_1) \\
 &= E^*(\text{trace}[\boldsymbol{\gamma}'_1 \mathbf{U}'_1 \mathbf{K}_j \mathbf{U}_1 \boldsymbol{\gamma}_1]) \\
 &= E^*(\text{trace}[\mathbf{U}'_1 \mathbf{K}_j \mathbf{U}_1 \boldsymbol{\gamma}_1 \boldsymbol{\gamma}'_1]) \\
 &= \text{trace}[\mathbf{U}'_1 \mathbf{K}_j \mathbf{U}_1 E^*(\boldsymbol{\gamma}_1 \boldsymbol{\gamma}'_1)] \\
 &= \text{trace}[\mathbf{U}'_1 \mathbf{K}_j \mathbf{U}_1 \mathbf{D}_1^{-1}] \\
 &= \text{trace}[\mathbf{K}_j \mathbf{U}_1 \mathbf{D}_1^{-1} \mathbf{U}'_1] \\
 &= \text{trace}[\mathbf{K}_j \mathbf{K}^+]
 \end{aligned}$$

where  $\mathbf{K}^+$  is the generalized (Moore-Penrose) inverse of  $\mathbf{K}$ . The relationships between weights  $w$  and expected penalties are complex and nonlinear. However, this expression for  $E^*(\boldsymbol{\pi}_j | w)$  allows us to quickly calculate the implications of a given set of weights, and therefore to construct a prior distribution for which all penalties have expected values that match their empirical averages in historical data.