

Fitting Spline Curves to LUV Ranges

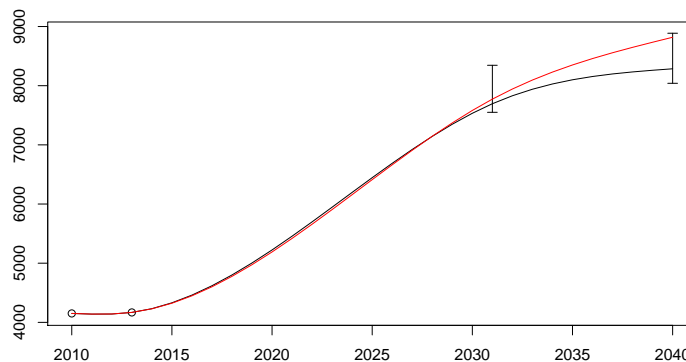
1 Data and Objectives

For each of the 148 jurisdictions, there are time series data on population and employment that include two observed data points, namely 2010, 2013, and two future estimates, namely 2040 (further called *second time estimate*) and one of 2025, 2030, 2031 or 2035 (further called *first time estimate*). Both future estimates are given as ranges, i.e. in form of a lower and upper bound.

The goal is to fit a spline such that it intersects exactly the observed data and is contained within the lower and upper bound of the future estimates. We define two criteria for the curves:

1. minimizing the curvature
2. maximizing the slope

For example, in the following figure, the red curve is preferable to the black curve, as it has less curvature and a higher slope.



2 Methodology

A simple methodology is used in which we create a large amount of possible splines and choose the one that fits our criteria best. Curvature is measured using the second derivative of the curve; for the slope the first derivative is used.

More specifically, the procedure consists of the following steps:

1. Randomly select n points between the lower and upper bound of the first time estimate, say a_1, \dots, a_n .
2. Randomly select n points between the lower and upper bound of the second time estimate, say b_1, \dots, b_n .
3. Randomly select R combinations of pairs $c_l = (a_i, b_j)$, where $R \leq n^2$, $l = 1, \dots, R$.
4. Attach each of c_1, \dots, c_R to observed data points and perform a natural spline interpolation for all remaining years. This results in R curves.
5. For each such curve remove data points up to 2013 and obtain the mean of the first derivatives on the remaining time points. Denote it by $d_1^{(1)}, \dots, d_R^{(1)}$.
6. For each curve remove data points up to 2013 and the last point of the curve, i.e. 2040, and obtain the mean of the second derivatives of the remaining time points. Denote it by $d_1^{(2)}, \dots, d_R^{(2)}$.
7. Since we want to maximize $d^{(1)}$ and minimize $d^{(2)}$, we select the curve for which
$$\max_{1 \leq i \leq R} \frac{d_i^{(1)}}{d_i^{(2)}}.$$

Setting $n = 100$ and $R = 1000$ yields reasonable results.

For questions on the methodology contact:

Hana Ševčíková

hsevcikova@psrc.org