

# Estimation of Contract Schedules using Constrained B-Splines

*Michael Chirico*

*February 18, 2017*

## Abstract

Most unionized teachers are paid according to a salary schedule (specifying wages as an increasing function of tenure and certification) explicated in contracts collectively bargained at the district level. With this in hand, teachers are able to infer their future potential wage trajectories at their own and other potential district employers. Lacking the physical contract faced by the teachers, an econometrician armed only with administrative data reporting actual wages in a given year must use some imputation techniques to deduce the wage structure. We explore the utility of natural Constrained B-Splines (COBS) to this end. COBS are an enhanced version of the traditional semiparametric splines technique enhanced by the ability to impose a monotonicity constraint on the resultant curve.

## Introduction

For many years, the ubiquitous characteristic of collectively bargained teachers' contracts has been the salary table, which gives a mapping from the calendar year, a teacher's experience (their length of tenure at the current district), and their certification (typically Master's vs. Bachelor's degree) to their wage. This table gives current teachers a clear understanding of how their pay will advance as a function of their labor inputs, and thereby gives forward-looking potential teachers and potential migrant teachers a clear understanding of their potential pay arcs under a district-switching decision-making framework, especially given that this information is typically openly available.

It would behoove an econometrician seeking to understand education labor market dynamics, then, to incorporate this information on future pay into their statistical modeling framework. Unfortunately, this data is typically not available in a format lending itself to easy analysis at scale – whether locked inside idiosyncratically formatted and sporadically-available contract PDFs or hidden behind large-scale freedom of information act inquiries, the temporal and financial costs of scraping such data into a usable form can be substantial.

Much more common in empirical settings is access to teacher-year-level salary data of the form  $y_{i,t} = y(\tau_{i,t}, c_{i,t}, d_{i,t}) + \varepsilon_{i,t}$ , where  $\tau_{i,t}$ ,  $c_{i,t}$  and  $d_{i,t}$  are the tenure, certification, and district of teacher  $i$  in year  $t$ , and  $\varepsilon_{i,t}$  represents unaccount factors affecting the wage (e.g., not all teachers work full time, and many teachers supplement their income with additional duties like coaching). The goal of this paper is to give one approach and some empirical lessons for trying to estimate the underlying mapping  $y(\tau, c, d)$  from such data.

## Method

There are a multitude of inference/imputation techniques suitable to the inference of a latent function of unknown parametric form available in the statistician/econometrician's palette. The powerful flexibility of nonparametric approaches (local regression, splines, Random Fourier Feature expansions) is a double-edged sword; as it happens, in this particular setting, even if we know linearity is not a reasonable functional form restriction, we do know some very basic properties of the underlying tenure-wage curves that will be violated in general by uninformed estimation techniques. In particular, we know that such tenure-wage curves are non-decreasing and that they are non-negative, i.e.,  $y(\tau', c, d) \geq y(\tau, c, d)$  whenever  $\tau' \geq \tau$ , and  $y(0, c, d) \geq 0$ .

He and Ng (1999) introduce a linear programming approach to incorporating monotonicity, curvature, and pointwise restraints to quantile regression spline estimation techniques, and Ng and Maechler (2007) present an overview of the R package `cobs` which gives an efficient implementation of this approach (COBS standing for Constrained B-Splines). The basic idea of quantile regression spline estimation is to swap out the standard squared loss function for a quantile-dependent weighted absolute loss function to target conditional quantiles instead of conditional means. Monotonicity, point, and curvature restrictions enter as penalized terms to the objective function; `cobs` expresses this in a fashion which facilitates the application of standard linear programming techniques for efficiency, and handles internally the issues of knot selection and penalty parameter assignment through cross-validation.

We implement and fine-tune this general approach with an eye to being as minimally-invasive as possible. The first innovation is required by the poor performance of standard COBS fit in extrapolation. Data sparsity in smaller districts means that it is often the case that only a small range of interesting  $\tau$  values are observed in a given year-certification-district. Monotonicity constraints are only built into the B-spline routine internally; the underlying basis functions may produce decreasing fits outside the observed range of data. To overcome this, we take a cue from the literature tackling Runge’s Phenomenon (Runge 1901), wherein polynomial approximations tend to exhibit extreme oscillations in extrapolation. This issue is one of the motivations behind natural cubic/smoothing splines (see, e.g., Friedman, Hastie, and Tibshirani 2001; Wahba 1990; Green and Silverman 1993; or de Boor 1978), which handle this issue by using a simple linear basis function outside the outermost interpolating knots. We incorporate this technique of linear extension only when necessary by testing the COBS fit for monotonicity;  $\tau$  values failing this constraint are replaced by extending the final non-decreasing fit values through the end of the range of extrapolation.

Next, a major shortcoming of COBS for this context is its limit to one-dimensional spline fits; while techniques for nonparametric B-spline fits are available in arbitrary dimensions (see de Boor 1978), at present COBS is only capable of imposing monotonicity on one dimension of a curve. In our context, however,  $y(\tau, c)$  is increasing not only with respect to  $\tau$ , but also with respect to  $c$  (as, without fail until only very recently in Wisconsin, certification was rewarded with a Master’s premium, typically a percentage increase in wage). One solution would be to generalize the implementation of COBS to handle a second dimension by simply adding penalty terms along this dimension. We abandon this approach because of the categorical nature of the certification dimension – there are not numerical units to the difference between having a Master’s vs. Bachelor’s degree; the assignment of such a number required by this approach would itself become an implementation hyperparameter, meaning that the ultimate fit would itself be sensitive to the choice of continuous representation.

Instead, we use a two-step procedure to fit the Bachelor’s and Master’s pay tracks in serial. In the first step, we fit the Bachelor’s career track as a typical one-dimensional COBS fit. In the second step, we first construct Master’s premia for each observation by subtracting out the predicted Bachelor’s pay corresponding to each observed level of tenure for a teacher with a Master’s degree. We then use COBS to fit a non-decreasing Master’s premium curve over all tenure levels on these residuals, before finally adding the Master’s premium and Bachelor’s fit curves to get the overall Master’s fit curve. Monotonicity of the result is guaranteed by forcing upwards monotonicity on the Master’s premium, a restriction in line with the empirical observation that Master’s degree pay is often simply a fixed-percentage rise over the corresponding Bachelor’s pay.

Our implementation was also aided by the imposition of concavity on the  $y(\tau, BA)$ . While not a theoretically-assured functional form restriction, concavity improves the goodness of fit notably. Small-sample district-level observations and simple reduced-form regressions of wages versus quadratic forms in experience support this shape’s validity. A variety of contracts obtained from a database for teachers in nearby Michigan also meet this condition, and the decrease in marginal returns to experience is also commonly found in the wider study of labor markets. We do not impose this restriction on the fit for the Master’s premium (the only restrictions there being non-negativity at 0 and, as mentioned, an increasing relationship with tenure).

## Data

The state of Wisconsin’s Department of Public Instruction (DPI) releases annual Salary, Position & Demographic reports through the WISEstaff data collection system, and these reports represent “a point-in-time collection of all staff members in public schools as of the 3rd Friday of September...” (Public Instruction 2017). The crucial data elements from each report are the Highest Degree Code (certification), Total Experience in Education (tenure), Total Salary, Total Fringe/Employee Benefits, and Agency of Work Location Code (unique identifier of district). We pull data from the 1994-95 academic year (AY) through AY2015-16 for a total of 3,588,614 teacher-position-year observations (many teachers serve in multiple roles within a school/district, and each of these is filed separately in the DPI system).

Data are first fed through the matching algorithm described in further detail in a companion paper. As a consequence, 26,304 (0.7%) observations are lost on account of belonging to teachers who could not be uniquely identified in a given year of data due to sharing a first name, last name, and birth year with another teacher in the data. Specific to the exercise at hand, with data reliability and precision in mind, we make a series of further restrictions on the data, as depicted in Figure 1.

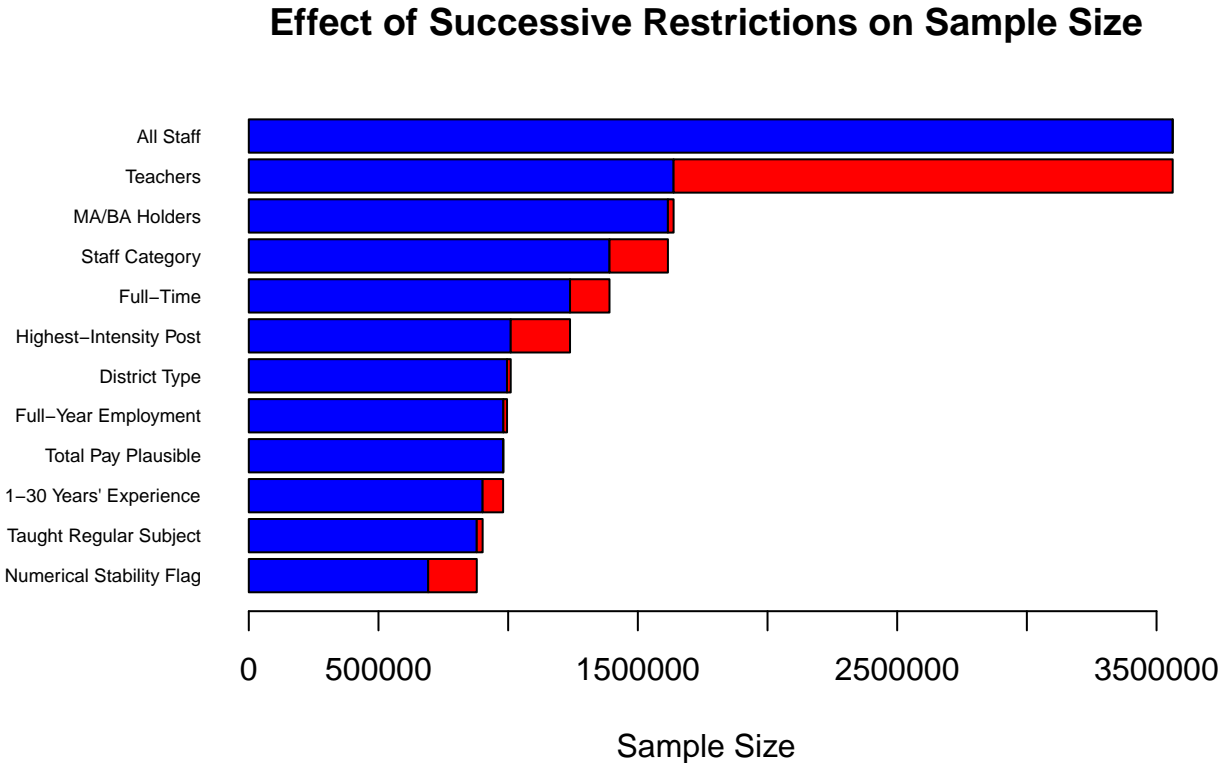


Figure 1: Sample Restrictions

To wit, from the full staff file are first eliminated all non-teaching positions; most common among the dropped observations are “Other Support Staff”, “Program Aide”, and “Short-term Substitute Teacher,” all of whom are likely to have their pay structured in ways that differ substantially from teachers (e.g., these positions are often part of a separately-bargained contract). We then remove all teachers whose highest degree is neither a Bachelor’s nor a Master’s degree, a minor restriction which allows substantial clarity in understanding districts’ pay structure. For more precision, we next eliminate teachers not categorized as professionals in regular education (as opposed by and large to those in special education)<sup>1</sup>.

The next two restrictions are related to a teacher’s full-time equivalency. Full-time equivalency is for the most part used to break down the relative use of staff’s time on various duties (for example, non-teaching time),

<sup>1</sup>Sensitivity analyses to this and other restrictions are given in the Appendix.

or to place substitute teachers on a pay grade commensurate with their work load. Given the possibility of split duties to dilute the contribution of a teacher’s classroom role to their pay, we eliminate teachers whose full-time equivalency across teaching roles is not 100; we then select only the highest-intensity (highest-FTE) role for each teacher in a given year<sup>2</sup>.

The next three restrictions have only a minor effect on sample size, and they are to eliminate districts not categorized as Wisconsin Public Schools (for the most part, observations failing this test are teachers employed by CESAs, Wisconsin’s supra-district administrative unit), to eliminate teachers recorded as having worked a period of time different from a full school year, and to eliminate a small number of teachers whose pay was implausibly small for a full time staff member (less than \$10,000 in a year). After this, we cut teachers outside 1-30 years’ experience for clarity and precision (sample sizes begin to drop off around 25 years of teaching), after which we eliminate teachers assigned to non-teaching subjects (health, academic support, gifted & talented support, etc.).

Finally, we make a series of cuts which are either required for COBS to function, or else increase the reliability of its output substantially. The most noteworthy/far-reaching of these numerical restrictions is to eliminate any teachers working in districts where there are not at least 20 total teachers in each degree track. While ultimately arbitrary, this number is reasonable to limit the potential effect of an individual teacher given the function to be fit to 30 levels of experience with minimal functional form restrictions. The other numerical flags require both the BA & MA track to be represented at a district, for at least 7 different levels of experience to be represented within a degree track, and for at least 5 unique values of the two measures of pay (salary and fringe benefits) to be available in each degree track; any teacher at a district failing these tests is dropped.

The sum total of all of these restrictions leaves us with an analysis sample of 691,591 teacher-year observations, made up of 94,495 individual teachers in 242 districts over 22 years. In total, there is sufficient data to fit 7,478  $y_t(\tau, c, d)$  curves, an average of roughly 100 observations per curve.

As an illustrative example of the patterns in the data we wish to quantify and formalize, we turn briefly now to Milwaukee Public Schools, the largest district in Wisconsin. Figure 2 depicts key moments of the empirical distribution of salary in 4 years at Milwaukee Public Schools, broken down by tenure and certification. The central lines on each plot (Bachelor’s pay track in blue, Master’s pay track in red) are the empirical median levels of pay, and thus give a rough approximation to  $y(\tau, c)$ . The dashed-line intervals on either side represent the 25th and 75th percentiles.

Notably, these intervals and the medians themselves tend to get quite noisy at later stages in the career, especially for the Bachelor’s track. This fact that reflects the almost universal certification of teachers by about 15 years into their career. This is reflected in the bar graph below each set of curves, which shows the distribution of teachers in each certification track by tenure. Almost all new teachers start with only a Bachelor’s degree; the relative presence of Master’s degrees grows over time as more teachers certify mid-career.

We can also note two more key empirical facts from this plot. First, the vanishing presence of teachers in both certification tracks leads the empirical median to be a poor approximation of  $y(\tau, c)$  since it frequently fails to respect the fundamental monotonicity constraint discussed above. With respect to tenure, this tends to affect the Bachelor’s track later in the career as more teachers certify, and the Master’s track very early in the career before teachers certify. The monotonicity with respect to  $c$  is maintained here for Milwaukee, but this is not always the case; our estimation procedure is thus careful to impose these restrictions internally.

Second, structural breaks are an important empirical phenomenon in this context. Each time a contract is renegotiated at a district, the tenure-wage curves can potentially change shape dramatically. It is with this in mind that we refrain, given our ignorance with respect to when such structural breaks occur, from combining information from adjacent years in fitting a given year’s curve, an approach which would substantially enhance the statistical power available to sparsely-populated districts. Such a structural break is apparent in Milwaukee, for example, between 2003 and 2004 and between 2005 and 2006, where the shape of the Master’s

---

<sup>2</sup>With a very small number of likely-erroneous exceptions, each teacher’s observation at a given district in a given year is associated with exactly one value for their pay (so that pay is not broken down by position), which means that neither can we “scale up” less-than-full-time teachers’ pay to full time values, nor is it particularly important to select the highest-intensity role for each teacher.

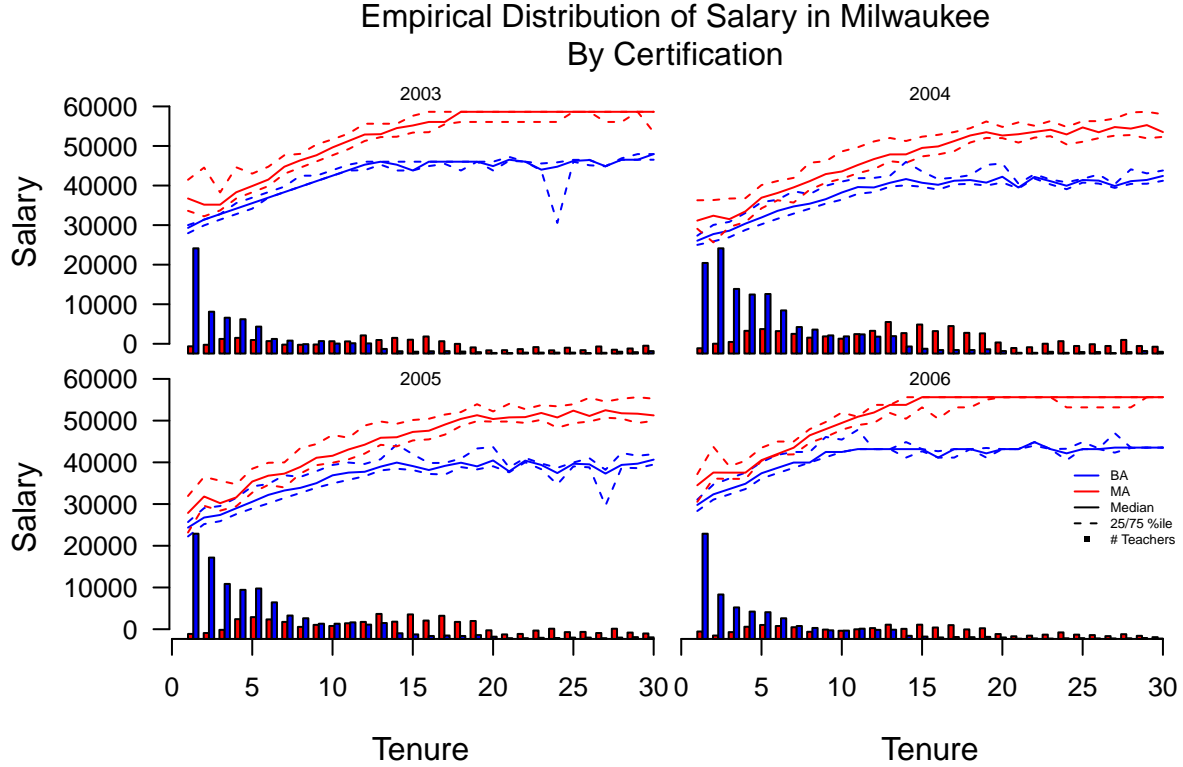


Figure 2: Pay in Milwaukee, 2003-2006

pay scale has shifted notably. While we eschew, for example, full Bayesian estimation of structural breaks in a given district, such techniques are applicable; instead, we settle for the purposes at hand with investigating evidence for such breaks *ex post*.

## Results

### Goodness of Fit

Returning to the motivating example illustrated in Figure 2, we turn first to the performance in Milwaukee, where, given the relative sample size, performance is expected to be very good. Indeed this is the case, as seen in Figure 3. The COBS fit has retained all the salient features of the empirical median return to experience and certification, while simultaneously improving over this nonparametric conditional median by ironing out nonmonotonicities found empirically as a result of small-sample bias.

Perhaps more telling is the goodness of fit in minimally small districts. Four such examples are featured in Figure 4. These four districts just barely satisfy the sample restriction that at least 20 teachers be present in both the BA and MA pay track (Melrose-Mindoro and Shell Lake have exactly 20 in each); for this reason, rather than plot the empirical median, we simply present the full distribution of wage and experience in these districts. Here again the COBS fit captures the essence of the wage-tenure curve even in these sparsely-staffed districts. While one can imagine Shell Lake or Waterford Union's actual contract schedule to feature the sort of late-career leveling that we have seen elsewhere, the power to identify and separate such features distinctly from those captured is strongly limited by the lack of data.

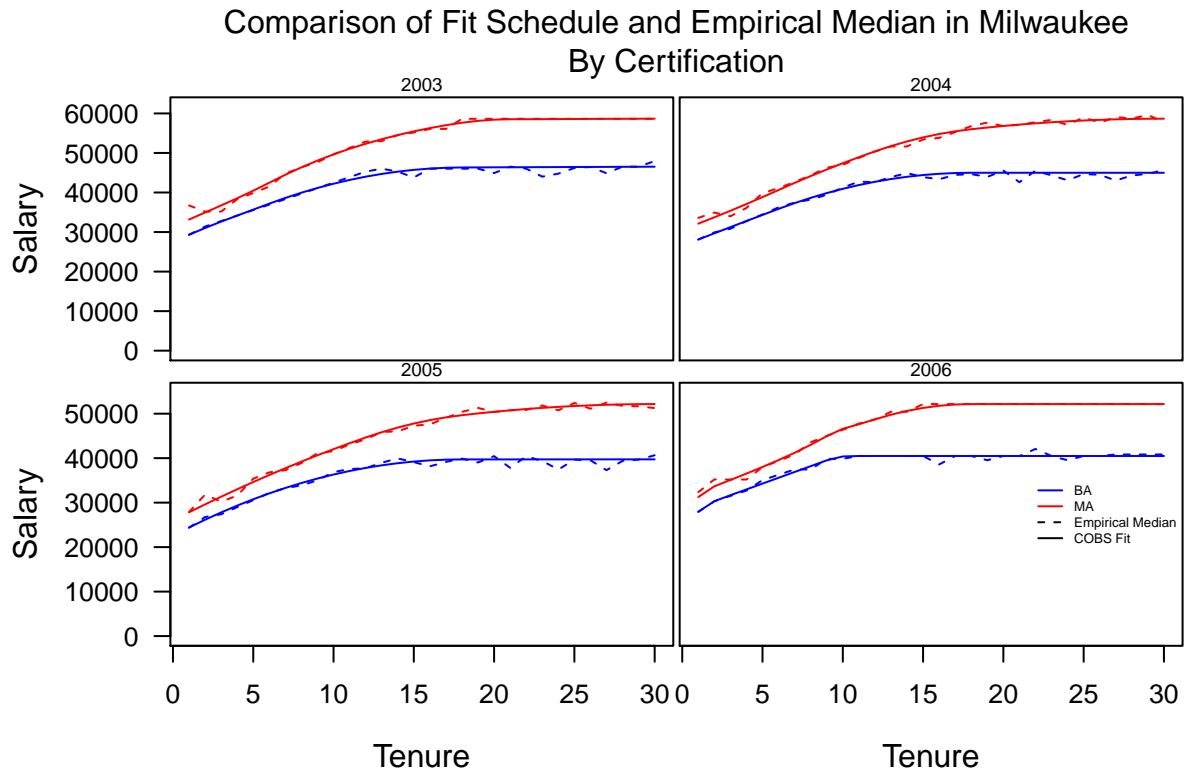


Figure 3: Estimation Results for Milwaukee, 2003-2006

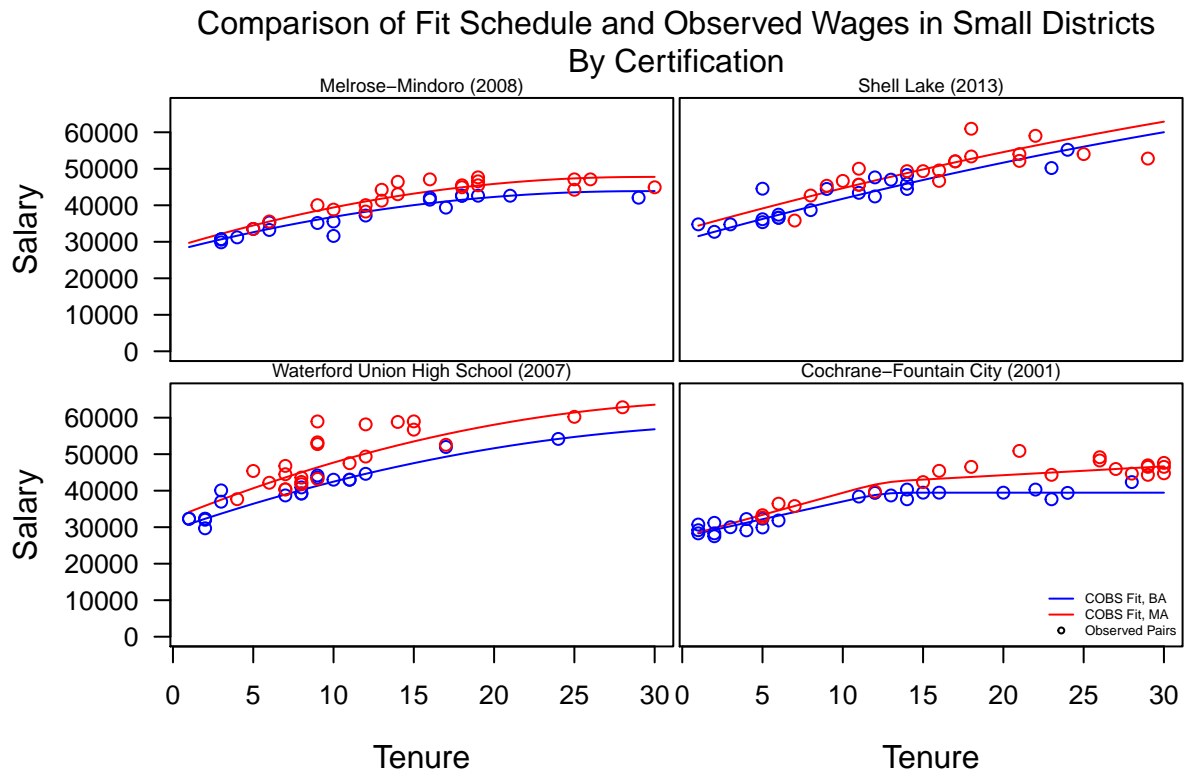


Figure 4: Estimation Results for Sparse Districts

# Characteristics of Wage Contracts in Wisconsin

## Conclusion

## References

- de Boor, Carl. 1978. *A Practical Guide to Splines*. Vol. 27. Springer-Verlag New York.
- Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2001. *The Elements of Statistical Learning*. Vol. 1. Springer.
- Green, Peter J, and Bernard W Silverman. 1993. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. CRC Press.
- He, Xuming, and Pin Ng. 1999. “COBS: Qualitatively Constrained Smoothing via Linear Programming.” *Computational Statistics* 14 (3). Springer: 315–38.
- Ng, Pin, and Martin Maechler. 2007. “A Fast and Efficient Implementation of Qualitatively Constrained Quantile Smoothing Splines.” *Statistical Modelling* 7 (4). SAGE: 315–28.
- Public Instruction, Wisconsin Department of. 2017. “School Staff: Salary, Position & Demographic Reports.” <https://dpi.wi.gov/cst/data-collections/staff/published-data>.
- Runge, Carl. 1901. “Über Empirische Funktionen Und Die Interpolation Zwischen Äquidistanten Ordinaten.” *Zeitschrift Für Mathematik Und Physik* 46 (224-243): 20.
- Wahba, Grace. 1990. *Spline Models for Observational Data*. SIAM.