University of Colorado Denver

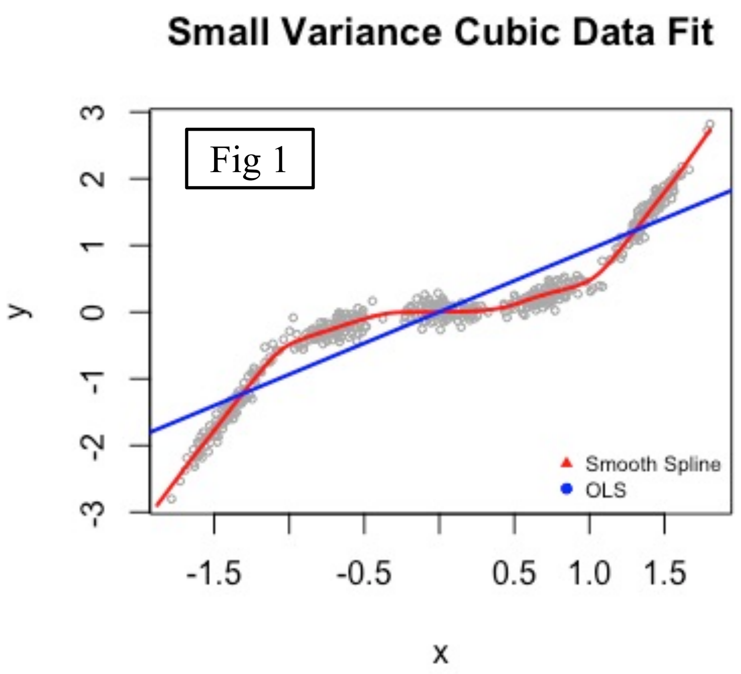
Investigating the properties and behavior of smoothing splines

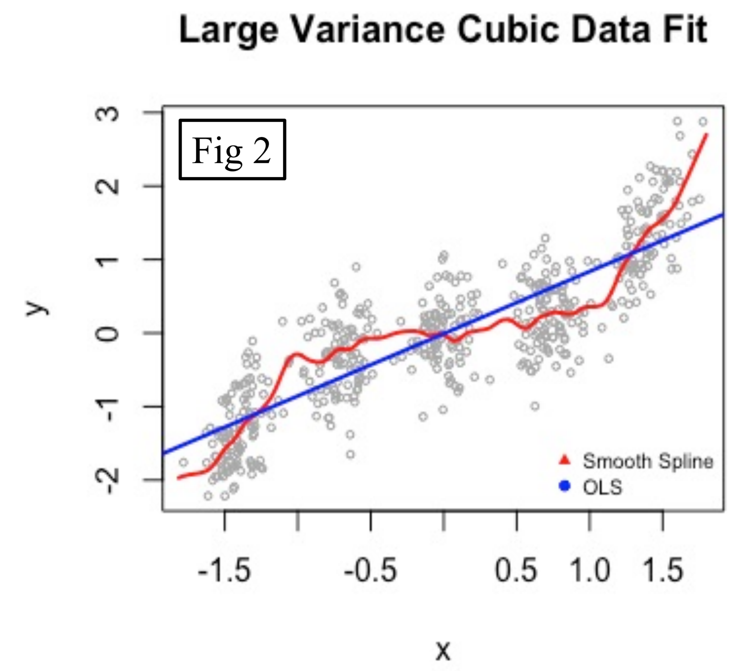
Lee Panter
University of Colorado-Denver
Mathematical and Statistical Sciences

Introduction

- Interpolation Splines
  - Piecewise polynomial function
  - Can be defined with varying degrees of continuity
  - Common to define "smooth" splines (makes interpolation more reasonable)
- Regression Splines
  - Combines the concepts of polynomial regression and step functions
  - Divide predictor space X into K partition
  - Fit a polynomial regression to the data within each subinterval
  - Restrict the fitted polynomials to preserve continuity at regional boundaries (Knots)
- Smoothing Splines
  - Result of a generalized RSS minimization problem
  - Penalized regression on variability in "solution function"
  - Piecewise cubic polynomial with knots at unique predictor space values
  - Continuous 1st and 2nd order derivatives at knot values

Motivation





Smoothing Spline and OLS fits are demonstrated for data that differs only in the induced response error variation.
Fig 1 unscaled induced response error=1
Fig 2 unscaled induced response error=5

MSE Calculations

Method	Small Variance Data MSE	Large Variance Data MSE
OLS	0.0680	0.14867
Smoothing Spline	0.0118	0.2016

Table 1

The MSE calculations illustrate the applicability of splines, but also how they could be potentially ill-suited for certain modeling situations

Smoothing Splines

- The function g(x) that minimizes the value of:

$$\sum_{i=1}^{n}\left(y_i - g(x_i)\right)^2 + \lambda \int g''(t)dt$$

- Where λ is a tuning parameter is known as a smoothing spline.
- For $\lambda = 0$ the function g(x) will interpolate the fitting data set, and as $\lambda \to \infty$ g(x)=OLS

- g(x) satisfies:
  - Cubic regression spline
  - Knots at each unique predictor value
  - Continuous 1st & 2nd derivatives
  - Linear in regions outside of extreme knots
- For a given λ we can write:
  $\hat{g}_\lambda = S_\lambda\, y$
  - $S_\lambda = \left(G^T G + \lambda\Omega_g\right)^{-1} G^T$
  - $G_{ij} = g_j(x_i)\ \ i,j = 1, \ldots, n$ is a basis matrix
  - $\Omega_g = \int g_i''(t) g_j''(t)dt$ is a penalty term
- $df_\lambda = \sum_{i=1}^{n}\{S_\lambda\}_{ii}$

- Use LOOCV to find λ

Methods

1. Create eight different data sets resembling data with a cubic trend, each with distinct values of induced response error from $\sigma^2 = 0.25$ to $\sigma^2 = 10$
   - Single predictor variable
   - N observations for each data set
2. Scale each data set to remove confounding effects of magnitude
3. Create K=5 cross-fold validation training and test sets from each data set
4. Create OLS and Smoothing Spline models for each data set using the corresponding training data
5. Use the created models to calculate predicted values of test data for each data set
6. Calculate MSE for each model on each data set
7. Calculate Mean MSE (MMSE) across K=5 fold values and evaluate the change in this value as σ^2 chages
8. Repeat steps (1)-(7) for N=50, 250, 500, and 1000 to determine if sample size is correlated with any effects discovered.

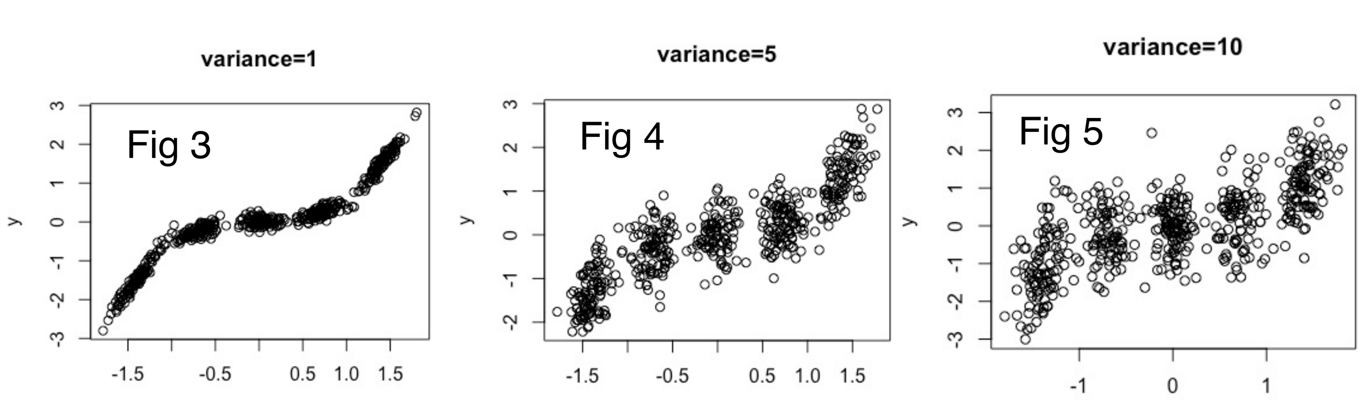
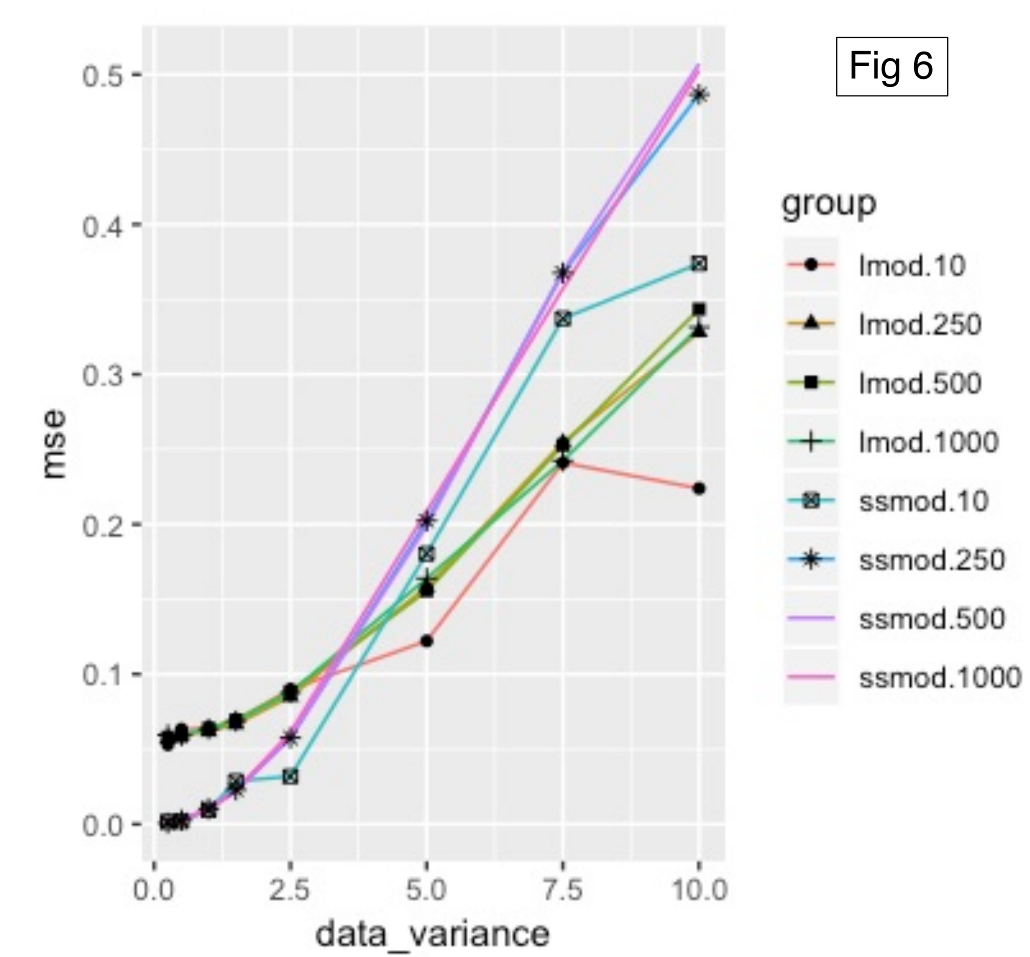


Fig (3)-Fig (5) Represent three of the 400 distinct data sets used in this methodology. They are test sets from the N=500 subset.

Results

Table 2

OLS Models				
variance	N=10	N=250	N=500	N=1000
0.25	0.0530	0.0570	0.0580	0.0595
0.5	0.0634	0.0595	0.0588	0.0592
1	0.0639	0.0618	0.0640	0.0635
1.5	0.0691	0.0666	0.0695	0.0688
2.5	0.0899	0.0851	0.0885	0.0876
5	0.1221	0.1583	0.1556	0.1639
7.5	0.2411	0.2545	0.2527	0.2424
10	0.2239	0.3283	0.3435	0.3315
Smoothing Splines				
variance	N=10	N=250	N=500	N=1000
0.25	0.0017	0.0006	0.0008	0.0008
0.5	0.0020	0.0027	0.0026	0.0026
1	0.0095	0.0103	0.0105	0.0100
1.5	0.0287	0.0229	0.0218	0.0222
2.5	0.0319	0.0576	0.0576	0.0612
5	0.1804	0.2023	0.1989	0.2085
7.5	0.3373	0.3676	0.3686	0.3571
10	0.3739	0.4867	0.5071	0.5029



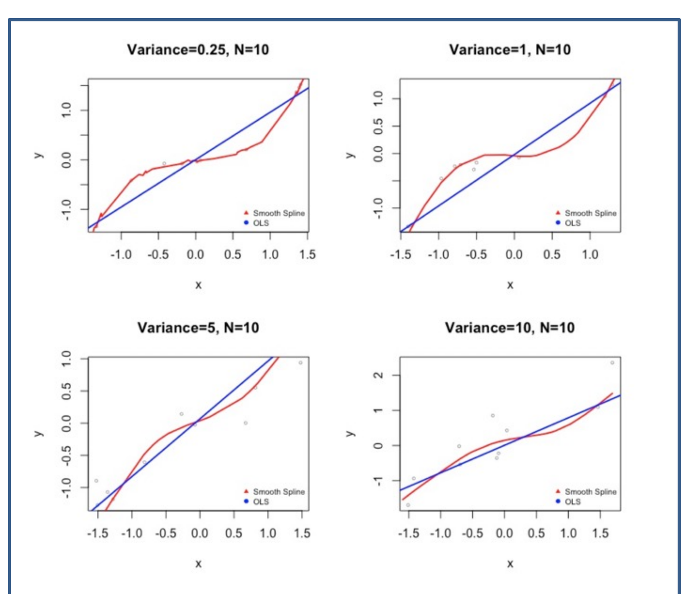


Fig7: N=10

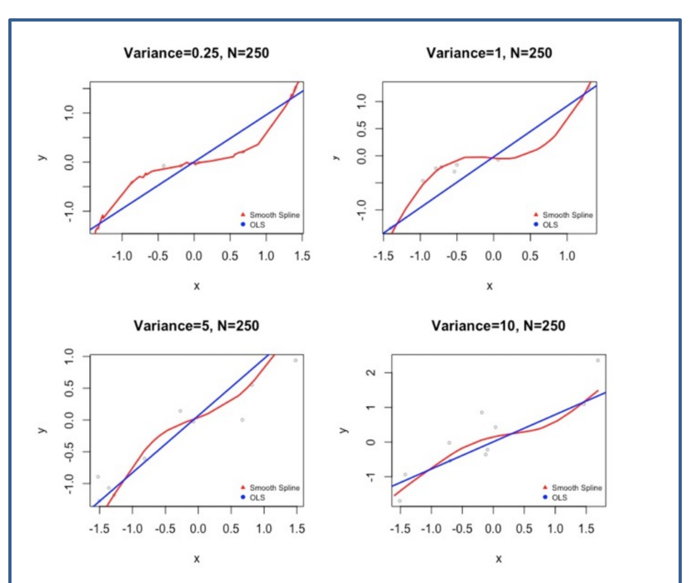


Fig7: N=250

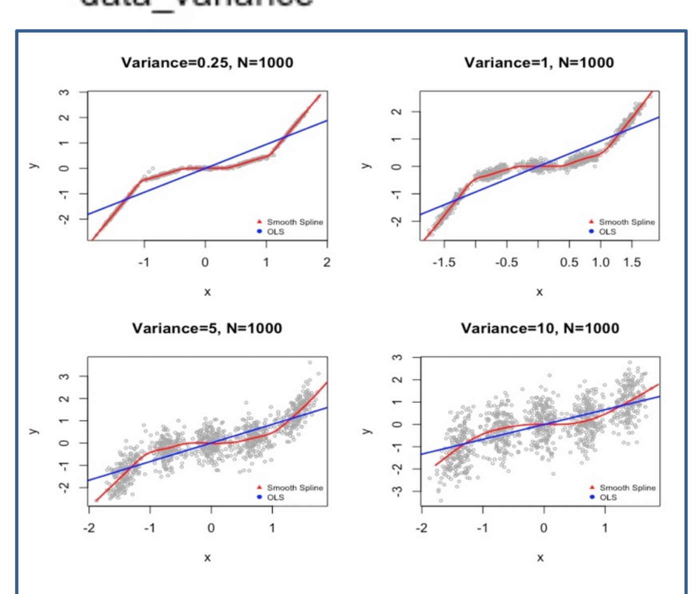


Fig7: N=1000

Conclusions and Future Research

- Initial Conclusions
  - Smoothing Splines are an excellent tool for providing very good models for data that has an underlying relationship with a polynomial, or any mapping that can be well-approximated using a polynomial.
  - Even though the theoretical convergence of a Smoothing Spline is to that of the OLS model, highly variable response values can lead to a less than optimal fit when compared to OLS fitting
  - Unless the underlying correlation between predictor(s) and response is very high, a Smoothing Spline model should be compared against OLS (among other methods)
- Future research interests
  - Exploring initial conditioning sensitivities, in particular with respect to influential observations, outliers, sampling methodologies and error distribution
  - Behavioral considerations in classification applications

References

- Geyer, Charle J. "5601 Notes: Smoothing." *Stat.umn.edu*, 2013, www.stat.umn.edu/geyer/5601/notes/smoo.pdf.
- Hastie, T., Tibshirani, R., & Jerome, F. (2017). *The Elements of Statistical Learning.* Springer.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An Introduction to Statistical Learning.* Springer.
- Tibs, r. (2014). *Smoothing Splines.* Retrieved from stat.cmu.edu: http://www.stat.cmu.edu/~ryantibs/advmethods/notes/smoothspline.pdf

For More Information

lee.panter@ucdenver.edu