Assessing the German Federal Budget with Distributional Methods

Christian Breunig

University of Washington

Working Paper June 2003

The purpose of this paper is to illustrate improvements for visually and statistically assessing budget change data. In particular, it will be shown that budget change data can be estimated via distribution functions, histograms, smoother density estimates, nonparametric regression, and relative distributions. The variable under investigation here is the difference in log of real spending by year and by governmental sub-function for the German Federal budgets between 1963-1989. The data set was provided Kathleen Bawn who acquired it through the Federal Ministry of Finance's Yearly Financial Report (*Finanzbericht*) (Bawn 1999: 723).

By describing budget changes more accurately, this paper contributes assessing evidence regarding the theory of punctuated equilibrium (Baumgartner and Jones 1993, Jones, Baumgartner, and True 1998, Baumgartner and Jones 2002). The notion of punctuated equilibrium purports that policy change occurs in an episodic and disjointed fashion. Hence, the model of policy change predicts that frequency distributions of outputs should display "'fat tails (indicative of punctuations), sharp central peaks (indicative of temporal stability), and 'weak shoulders' (indicative of a relative lack of moderate change)" (Jones and Sulkin 2002: 11). Thus, frequency distributions should be leptokurtic. In order to determine the degree of leptokurtosis, the actual budget data is compared to a hypothetical Gaussian distribution with the same mean and variance.

Before examining the budget distribution in more detail, Table 1 (see Appendix[1]) presents some basic descriptive statistics[2]. Figure 1 displays the most fundamental assessment of the German data as well as a hypothetical Gaussian distribution. The cumulative density function (CDF)[3] including its 95% confidence bounds basically gives the probability that a randomly

[1] I employed the statistical package R here.
[2] I decided to drop one data point from the sample because its value suggested that it might be a faulty data entry.
[3] Throughout the paper, it is assumed that budget change is a continuous random variable *X* that is independent and identically distributed from the distribution *F*.

chosen values is less than or equal to *x*. Figure 1 very shows that the budget data clearly has weak shoulders. For example, one can discern that the difference in proportions of German and Gaussian distribution at 30 percent change is approximately 21 percent points. Second, the graph hints to the existence of “fat” tails because the budget line crosses the Gaussian at roughly the 100 percent change level. Third, the steep nature of the empirical slope also suggests that a high amount of stability is present at the center of the distribution. Figure 2 – the quantile function – confirms the findings of Figure 1. In general, the quantile $Q(p)$ can be defined as the value of y below which a proportion of p of the value falls. The weakness of the shoulders is illustrated by the fact that the difference in the German budget changes and a Gaussian distribution at the 75^{th} percentile is -20.86 percent points. In other words, at the 75^{th} percentile, the German budget change is roughly 21 percent points below the expectation derived from the Normal curve.

Previous studies relied heavily on the comparison between a Gaussian distribution and a histogram of budget change. John and Margetts (2001: 21) suggest that the bin width is more or less arbitrarily chosen and that the choice of bin width has considerable impact on evidence regarding leptokurtosis. Figures 3 and 4 consider and alleviate their claim. Figure 3 portrays the histogram of German budget changes using the default bin width chosen by the statistical package R overlaid with a Gaussian distribution. On first glance, it appears that the Gaussian distribution does a fair job in representing the empirical distribution. However, given that the 884 data points are grouped in just a few categories and that the tick marks for budget change on top of the distribution are spread out more widely, using the default bin width may not accurately reflect the data. Sheather and Jones (1991) as well as Simonoff (1996) shows that a trade of between bias and variance exist when choosing the bin width and that it is possible to minimize this tradeoff. The optimal bin width *h* for the German budget data is roughly 8.59 percent points, which corresponds to 89 bins for the data set. Figure 4 displays the German budget data using the optimal bandwidth. When using the optimal bin width, Figure 4 portrays that – compared to the Normal curve – the budget distribution is clearly leptokurtic in nature.

In general, a histogram is easy to interpret and construct but – as a step function – it does not adapt to the relative density shape. In other words, the general drawback of histograms for estimating univariate densities (PDF) is that they are step functions, while the data is usually continuous and much smoother. This suggests that alternative estimators exist that may better reflect the properties of budget change data. In this paper, two smoother methods that assess local characteristics are utilized for estimating the probability density function (PDF). First, a kernel density estimator (KDE) does reduce the bias from h to h^2 and thus detects local attributes better than a histogram. However, a KDE possesses two disadvantages: the lack of local variation in smoothing and the flattening of peaks and valleys. Second, local likelihood density estimates (LLDE) might provide an improvement over a KDE because a LLDE has well known properties (such as standard error) and allows choosing the bandwidth adaptively[4].

Figure 5 displays the PDF of German budget data using the default kernel density estimator and Sheather Jones optimal as well as half and twice of the optimal bandwidth. The lines in Figure 5 show that the KDE identifies local characteristics of the budget distribution and that the optimal Sheather Jones bandwidth is sufficient for detecting these characteristics. When compared to the Gaussian distribution (blue line), all three KDEs nicely show that the German budget is distributed in leptokurtic fashion. The KDE identifies an unusually high peak at the center, weak shoulders, and “bumps” on the tails. Figure 6 confirms the basic findings of Figure 5. The estimates in Figure 6 are obtained by using local likelihood methods (LLDE), which

[4] However, LLDE assuming that the relative density is actually member of a certain family of distributions.

assess 10% and 35% of the neighborhood around any given point. A visual comparison between the estimates displayed in Figures 5 and 6 suggests that both methods do roughly equally well.

Jones and Sulkin (2002) suggest that regression model can also be employed for detecting the leptokurtic nature of distributions. For Figure 7 and 8, I employed local nonparametric regression in order to regress German budget changes on Gaussian change. Again, this method allows for taking local features of the data set into account. Figure 7 employs a locally weighted scatterplot smoother (*lowess*). *Lowess* computes the estimate for each data point by assessing and weighting a point for a window of 10% and 50% percent of the data. The *loess* function used for generating Figure 8 proceeds similar to *lowess* but uses a different weighting methods and can be employed for multivariate analysis. Both methods produce the same basic results. Figure 7 and 8 clearly show that the German budget data is not normal. If the data would be normal, the data points would be aligned along the linear prediction. Instead, the *lowess* and *loess* estimates show that fewer than expected medium range changes exist and than the tails for the empirical distributions are much stronger than the tails of the Gaussian[5]. Moreover, the different estimates for a 10% and 50% neighborhood suggest that the more local features of the budget distribution are taken into account the more the budget distribution differs from the Normal distribution. Overall, using nonparametric regression methods appear to be a powerful tool for delineating budget change.

The last methods for comparing two distributions considered here are relative distributions (Handcock and Morris 1999). Relative distributional methods combine previously considered methods, require only weak assumptions about the distributions, and are invariant to

[5] I tried to obtain a measure of kurtosis from the comparison between the linear and the lowess estimate. Kurtosis in this case would be defined as the area between the linear and the lowess prediction. Thus, a kurtosis measure would be obtained by integrating the area of two functions from the minimum to the maximum value of the budget data. Unfortunately, I was unable to write the code for doing this.

measurement scale. Figure 9 and 10 show the relative distribution of Gaussian distribution (reference) to German budget changes (comparison) using different bin width. The interpretation of the Figures 9 and 10 is as follows: the relative density is the ratio of the frequency of the comparison distribution (German budget changes) to the frequency of the reference distribution (Gaussian) at the r^{th} quantile of the reference distribution level. Again, for the estimates, different degrees of local features will be considered (10%, 35%, and 70% of the distribution for each point).

The density estimates provide an intuitive assessment of the two distributions. If both distributions would be distribution Gaussian then the relative distributions would be uniform. These graph makes clear that the biggest difference between the two distributions is at the center of the distribution: more than two times as many budget changes as expected by the Gaussian distribution are in the fourth decile defined by the Gaussian distribution. The fifth and sixth decile is more then 1.5 times higher as expected. Both tails – the two lowest and the two highest deciles – have about half as many budget changes as expected by the Gaussian distribution. Figure 9 may be problematic for detecting extreme changes because the bottom and top deciles are summarize with one each bar. Figure 10 improves this weakness. I plotted the relative density with 100 bins in order to visually check whether a high amount of large changes can be detected. Although the overall interpretability suffers by plotting 100 bins, Figure 10 shows that in the top one percentile empirical budget changes occurred more often than the normality assumption would expect. In addition, the local estimators have a positive slope a each tail. This also indicates that change is more extreme at the tails of the German budget distribution.

In summary, the data presented in this paper confirms the prediction that budgets are distributed leptokurtic, i.e. the punctuated equilibrium model can successfully describe policy change. Three findings of this paper should be considered for estimating budget distributions. First, bin width can and should be optimized in order to accurately display the investigated distribution. Second, using cumulative density and quantile functions, as well as densities based on kernel or local likelihood methods are indispensable for describing univariate distribution comprehensively. Finally, nonparametric regression methods and relative distributions provide intuitive assessments for comparisons between distributions. Regarding the German budgets, future analysis should consider differences in budget changes over time based on these methods.

Bibliography

Baumgartner, Frank and Bryan Jones. 1993. *Agendas and Instability in American Politics*. Chicago: University of Chicago Press.

Baumgartner, Frank and Bryan Jones. 2002. *Policy Dynamics*. Chicago: University of Chicago Press.

Bawn, Kathleen. 1999. Money and Majorities in the Federal Republic of Germany: Evidence for a Veto Players Model of Government Spending. *American Journal of Political Science* 43 (3): 707-736.

Handcock, Mark and Martina Morris. 1999. Relative Distribution Methods in the Social Sciences. New York: Springer.

John, Peter and Helen Margetts. 2002. *Policy Punctuations in the UK: Fluctuations and Equilibria in Central Government Expenditure since 1951*. Unpublished Paper.

Jones, Bryan D., Frank R. Baumgartner, and James L. True. 1998. Policy Punctuations: US Budget Authority, 1947-1995. *Journal of Politics*, 60: 1-33.

Jones, Bryan and Tracy Sulkin. 2002. *Policy Punctuations in American Political Institutions*. University of Washington. Working Paper.

R Development Core Team. 2003. *R: A Language and Environment for Statistical Computing*. http://www.R-project.org

Sheather, S. J. and Jones, M. C. 1991. "A Reliable Data-based Bandwidth Selection Method for Kernel Density Estimation". *Journal of the Royal Statistical Society series B* (53), 683-690.

Simonoff, Jeffrey 1996. *Smoothing Methods in Statistics.* New York: Springer.

Appendix

Table 1: Descriptive Statistics for German Budget Changes, 1963-1989.

Statistic	
Mean	12.55
Median	9.13
Std. Deviation	34.52
Minimum	-179.60
1st Qu.	1.62
3rd Qu.	19.34
Maximum	585.50
Interquartile Range	17.71
N	884

Fig. 1: Cumulative Distribution Function of Budget Changes

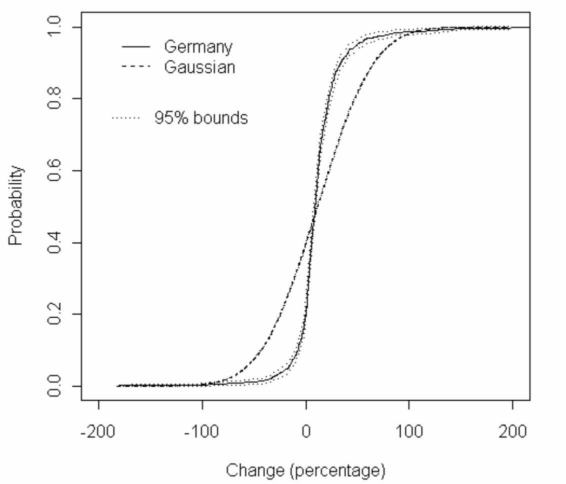

Fig 2: Quantile Function of the Budget Changes

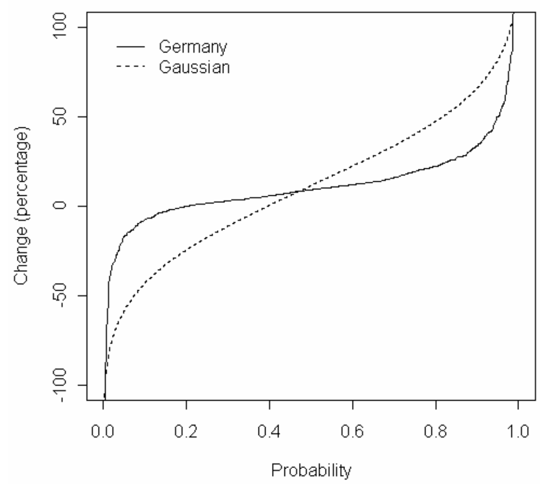

Fig 3: Histogram of Budget Changes with Default Bandwidth

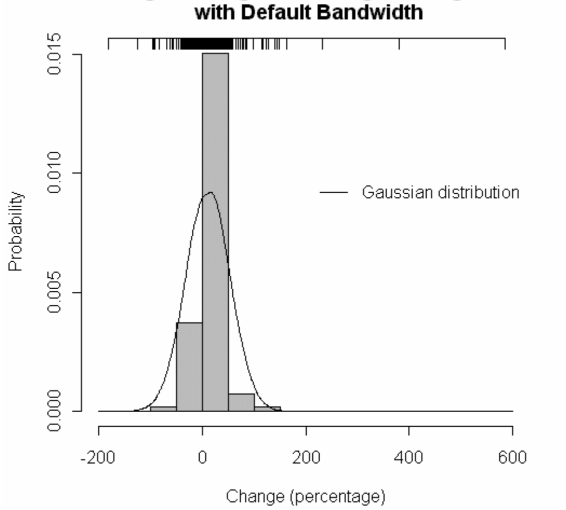

Fig 4: Histogram of Budget Changes with Optimal Bandwidth

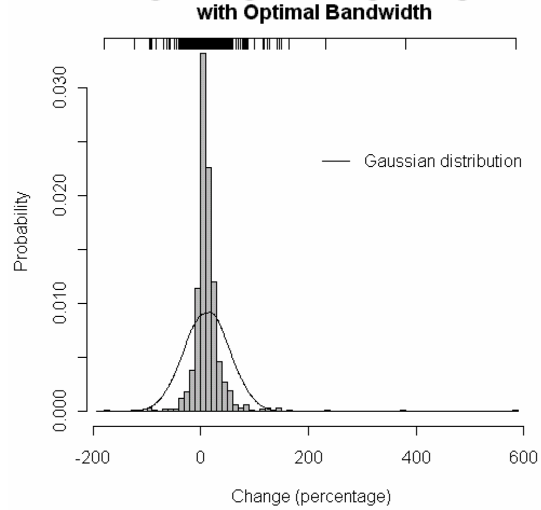

Fig 5: Histogram of Budget Changes

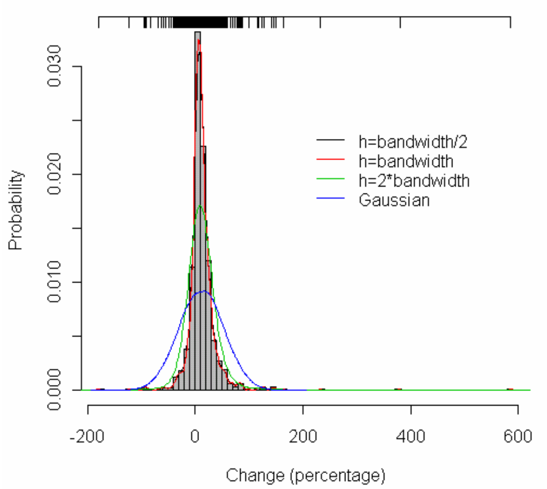

Fig 6: Histogram of Budget Changes

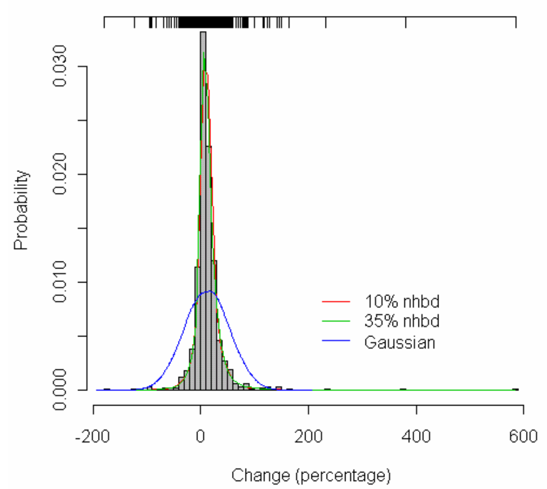

Fig 7: Scatterplot of Gaussian Curve to Budget Changes

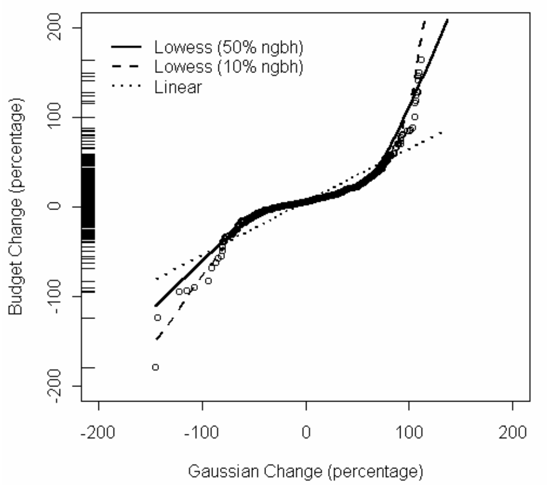

Fig 8: Scatterplot of Gaussian Curve to Budget Changes

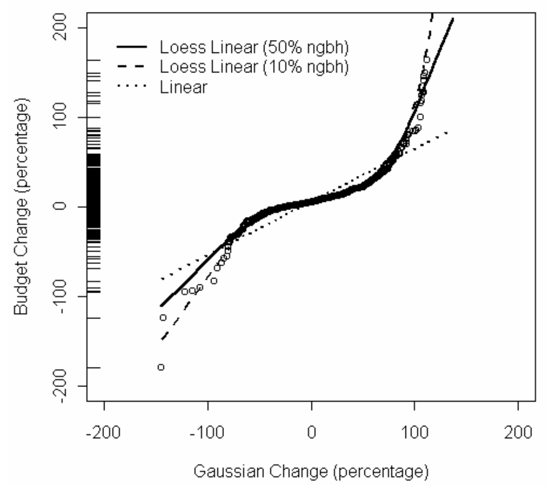

Fig 9: Relative Distribution

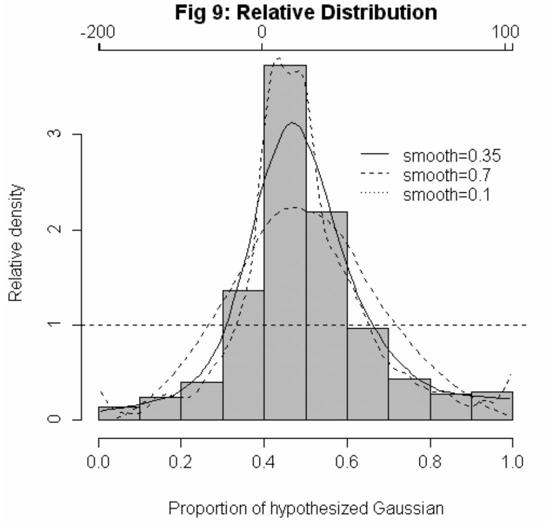

Fig 10: Relative Distribution with 100 bins