A Comparative Distributional Method for Public Administration

Illustrated using Public Budget Data

Jens Ledet Jensen (jlj@math.au.dk)[1], Peter B. Mortensen (peter@ps.au.dk)[2] and Soeren Serritzlew (soren@ps.au.dk)[2]

[1]Department of Mathematics, Aarhus University, Denmark

[2]Department of Political Science, Aarhus University, Denmark

Abstract

Many outputs and outcomes of relevance to public administration research are distributed in a way that escapes simple statistical descriptives. The field of public administration has generally been slow to acknowledge this fact, but in recent years the research field of public budgeting has experienced what we term a "distributional turn" reflected in a massive increase in publications studying distributions of changes in public budgets. Yet, distributional budget studies still rely on inadequate statistical methods. In this paper, we introduce a comparative distributional method (CDM), which allows for inferential statistics and statistical modeling of distributions. It also allows for specification of parameter values that precisely characterize the empirical distribution. All this is important for systematic comparisons of empirical distributions. We illustrate the usefulness of the CDM approach by analyzing budget data from seven countries and conclude with a discussion of how this new method can be extended to other areas of public administration than public budgeting.

When studying climate, researchers are not merely interested in average climate changes. The probability of extreme events, worst case scenarios, is at least as important as the mean or median changes. When offshore installations are designed, knowledge of the probability distributions for extreme wave and crest heights is at least as important to engineers as knowledge about the height of the average wave. When it comes to stock markets, it is normally not the average share prices that cause bankruptcy but the occurrence of extreme changes. What researchers strive for, in all these cases, are models that closely approximate the underlying probability distributions of small, medium and extreme events. Furthermore, in all these cases the standard well-behaved bell curve is not a very good approximation of the heavy-tailed probability distributions underlying these events.

The same is true of distributions of year-to-year changes in public budgets. The distributions of these budget change events are characterized by a high central peak, slender shoulders, and more extreme observations than expected from a normal distribution of similar mean and variance. This empirical generalization of public budget distributions holds across space (different countries and states), over time, and across government tiers (from the local to the federal level) (Jones et al. 2009). Public agencies and organizations normally do not go bankrupt if they are subject to an extreme budget change, but extreme budget changes may matter just as much, or more, to the performance of public organizations than average budget changes (O'Toole and Meier 2003, Andersen and Mortensen 2010).

Studying full distributions of budget changes, not just the annual mean change in a budget, informs us about one of the most important recurrent decisions in a political system. Knowing the probability of observing minor, medium and large-scale year-to-year budget changes gives us information about administrative and political decision-making more generally. That is exactly why political scientists for more than fifty years have been involved in a quest for better models of the distribution of small and large-scale changes in public budgets (Davis, Dempster and Wildavsky 1966, Davis, Dempster and Wildavsky 1974, Padgett 1980). In fact, in recent years, the scholarly interest in the shape of empirical distributions of year-to-year budget changes has been so intense that it looks as if there has been a "distributional turn" within the research field of public budgeting (see e.g. Jones and Baumgartner 2005a, Robinson et al. 2007, Jones et al. 2009, Breunig, Koski and Mortensen 2010, Breunig and Jones 2011, Jones, Zalanyi and Erdi 2014, Robinson, Flink and King 2014, Flink 2018). Many of these studies have been published in the *Journal of Public Administration Research and Theory*.

Despite these landmark contributions, better statistical tools are needed to fullfil the potential of distributional studies in public budgeting. In this paper, we propose a comparative distributional method (CDM) based on the normal inverse gaussian distribution (NIG) (Barndorff-Nielsen 1997). This methodology allows for inferential statistics, statistical modeling of budget distributions, and prediction. Inferential statistics makes it possible for the first time to compare budget distributions from different political systems and different policy areas to see if they are statistically separable. The statistical modelling allows for specification of parameter values that precisely characterize the distributions and predicting the likelihood of for instance observing a 10 percent cut in welfare spending.

Other areas of public administration research besides public budgeting have been slow to acknowledge the insights that can be gained from distributional analysis, but the CDM is relevant to many other topics. The 'big data' movement in public administration has increased the need for methods to analyze not only averages but also the distributions behind them. The study of public sector reform can also benefit from incorporating distributional analysis. To understand, for instance, how student performance is affected by a given school reform, it can be useful to have a precise estimation of the full distribution of student performance, not just the average effects. Such predictions can be very useful seen from the perspective of reformers and policy makers. In this paper, we focus on how CDM can advance the study of public budgets, but in the conclusion we return to the question of the broader relevance of CDM to public administration research.

In the next sections, we first review the distributional turn in the study of public budgeting. We then discuss some limitations of the standard approach to distributional analysis, and the advantages of adopting the comparative distributional method. In the empirical part we perform a comparative distributional analysis of more than 47,000 observations of budget changes covering national, state, and local budgeting from seven different countries. We demonstrate how the CDM can be used to calculate probabilities of exact events, estimate precise models for empirical distributions, estimate statistical significance, and confidence intervals for parameter estimates, and to test significance of differences between budget distributions.

1 Distributional analyses of public budgets

The first quantitative budget study by Wildavsky and colleagues focused on the mean of annual changes in public budgets and concluded that: "This year's budget is based on last year's budget, with special attention given to a narrow range of increases or decreases" (Davis, Dempster and Wildavsky 1966, 529). This model of budget changes can be written as:

$$B_t = B_{t-1} + e_t, \tag{1}$$

where B_t (B_{t-1}) is the budget in year t ($t-1$) and e_t is an error term assumed to be random and normally distributed (see Davis, Dempster and Wildavsky 1966, Davis, Dempster and Wildavsky 1974). By subtracting B_{t-1} from both sides of the equation, it can easily be seen that this equation implies that any year-to-year budget change is simply a random component. The normal distribution is a natural model for the budget change within a year, being the result of a number of decision steps and using the central limit theorem from probability theory.

Empirically, however, annual (percentage) changes in public budgets are not normally distributed. Padgett (1980) was the first to show this, using US domestic program allocation data from 1957, 1964, and 1966. Instead of a normal distribution, Padgett found that the distribution of annual changes in these data looked more like an exponential or a Pareto distribution, which are both characterized by more extreme observations than a normal distribution of equal mean and variance.

For many years, Padgett's distributional approach to public budgeting did not gain much scholarly attention (Bendor 2003). However, in the 2000s, Jones, Baumgartner and colleagues revitalized the interest in distributional budget analyses (see e.g. Jones, Sulkin and Larsen 2003, Jones and Baumgartner 2005a, Jones and Baumgartner 2005b, Jones et al. 2009, Jones, Zalanyi and Erdi 2014). In line with Padgett's findings, this research shows that all known budget time series follow a characteristic pattern: A distribution with a high peak, slender shoulders and fat tails.

This change in scholarly focus, what we term the distributional turn, led to the discovery of a striking empirical regularity in space, over time, and across government tiers (Jones et al. 2009). In budget literature jargon, such distributions are termed leptokurtic distributions, i.e., the kurtosis of these distributions is relatively high. Figure 1 shows a characteristic budget change frequency distribution based on annual percentage changes of subfunction budget authority in the US from 1947 to 2012. The distribution of budget changes is clearly not normal. Compared to the theoretical normal distribution with similar mean and variance, it has a strong central peak, weak shoulders and long, moderately fat tails.

Breunig and Jones (2011) provide an excellent review and illustration of the state of the art in distributional analyses of public budgeting research. Distributional analyses have often been based on two sets of tools, which we review below, namely descriptive statistics, based on the kurtosis value, and a distributional model, based on gaussian, exponential, or Paretean distributions. There are some important limitations with both of these two methods: Descriptive statistics are insufficient to characterize the form of a distribution, and standard distributional models do not allow formal statistical tests, precise modeling, or prediction.[1]

[1]Some scholars also apply quantile regressions, but given that this tool tells us nothing about the shape of the empirical distributions, it is not relevant to this discussion.

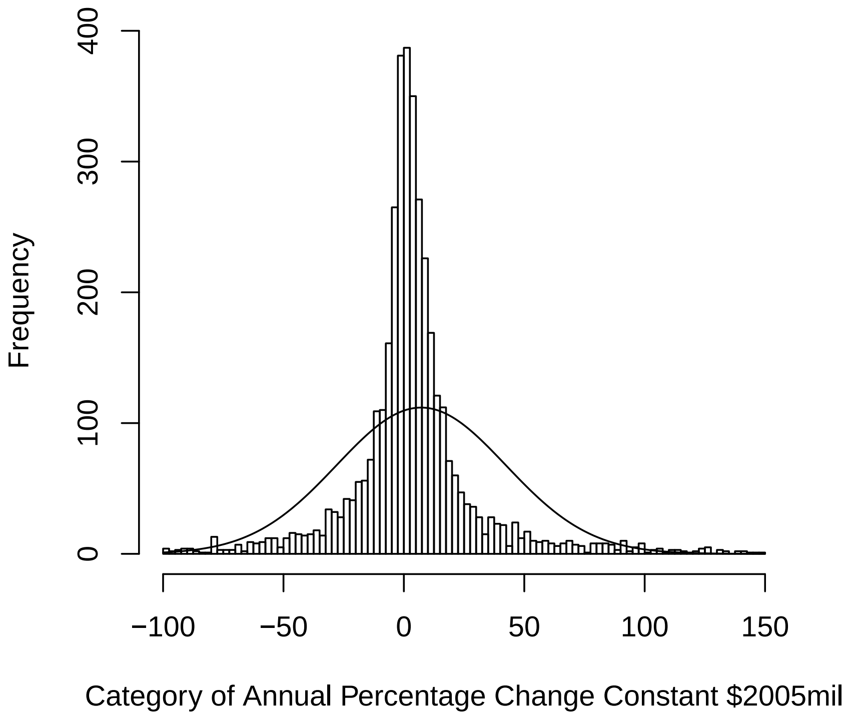

Figure 1: Distribution of annual percentage changes of subfunction in US Congressional budget authority from 1947-2012. The histogram has been cut at -100 and 150 for presentation (there are 8 instances below -100 and 96 above 150). Source:www.comparativeagendas.net.

Descriptive statistics

Kurtosis is the fourth central moment divided by the square of the variance and, in most common applications, subtracted the number three (excess kurtosis). This is a scale-free summary measure of the shape of the distribution with the normal distribution having the value zero. Kurtosis is an indicator of how tall and skinny or short and squat a distribution is compared to a normal distribution of similar mean and variance. The empirical kurtosis based on a set of data is known to be sensitive to extreme observations, and many scholars use L-kurtosis instead (Hosking 1990), which is a more stable measure using only first-order moments of order statistics.[2] A standard Gaussian distribution has an L-kurtosis score of 0.1226, whereas the L-kurtosis values of empirical budget distributions often

[2]See Supplementary Material SM B for an introduction to L-kurtosis.

range between 0.3 and 0.6.

The simple summary measure of kurtosis or L-kurtosis may tell us something about the distribution, but as noted by Breunig and Jones (2011, 108), they do not do justice to the theoretical interest in the full probability distributions. Differently shaped empirical distributions can have similar levels of kurtosis, which reflects that one parameter (the fourth standardized moment) provides only rough information about the shape of the distribution. Particularly, the measure of kurtosis does not adequately represent what may be important differences in the center, the peak, of the budget distributions.

The left side of Figure 2 shows a comparison between two distributions with the same value of the kurtosis. For the solid curve, the kurtosis of 2.08 is caused by a sufficiently high probability in the extreme tails of the distribution; for the dashed curve, the kurtosis is caused by a large probability close to zero. The right side of Figure 2 shows a comparison of two distributions with widely different kurtosis values. This difference is caused by the part of the distribution beyond ± 4 that contributes less than one percent to the total probability and that only become visible on a log scale for the density. A very similar picture emerges for L-kurtosis: Identical L-kurtosis can follow from very different distributions (as in the left panel of Figure 2), and very different L-kurtosis values can be inferred from distributions that seem to be very similar (as in the right panel of Figure 2). This illustrates the point that the kurtosis measure neither can nor has been designed to provide a complete description of the form of the distribution. Its usefullness in comparative distributional budget analyses is therefore limited.

Standard distributional models

Scholars normally do not exclusively rely on the simple summary measure of kurtosis or L-kurtosis when they investigate the distributions of budget changes. Another popular tool is goodness-of-fit tests that assess if the probability distribution underlying the budget observations could have come from a specified distribution (Breunig and Jones 2011). Given the special interest in the Normal distribution, scholars often conduct a test for normality using the Shapiro-Wilk test. Regarding the double exponential or double Pareto

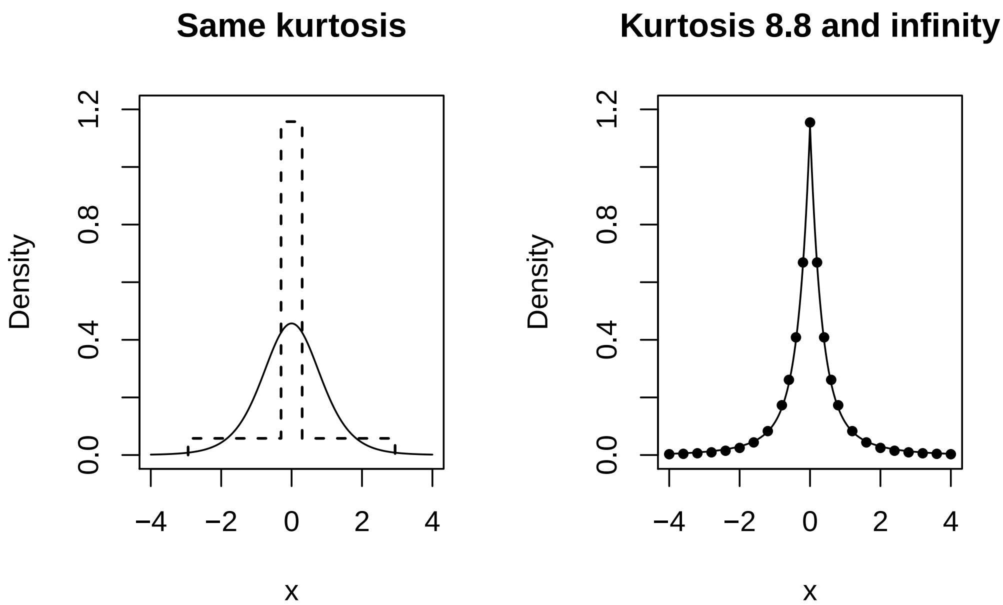

Figure 2: On the left is a comparison between a t distribution (solid) and a mixture of two uniform distributions (dashed), both having kurtosis 2.08 and variance one. On the right is a comparison of a double pareto (with $\beta = 4$) having infinite kurtosis (dots) and a mixture of two double exponentials with a kurtosis of 8.79 (solid).

distribution, the Kolmogorov-Smirnov test has been considered although p-values are not known and have to be found by simulations. The Kolmogorov-Smirnov test focuses on the center of the distribution and less on the tails.

The comparison of different types of density distributions addresses the limitations of the simple summary measures of kurtosis or L-kurtosis. However, the value of the comparison of density distributions is reduced by the categorical distinction between normal, exponential or Paretian. Figure 3 illustrates the large gap between a normal distribution on one side and a double exponential or the class of double Pareto on the other side. The double exponential has little flexibility, and the double Pareto models deviate from the normal distribution with very heavy tails only. This means that many empirical distributions are not well described by any of the three classes of distributions.

Many scholars supplement the goodness-of-fit tests with visual inspections of histograms of budget changes or plots of cumulative tail probabilities in order to get an

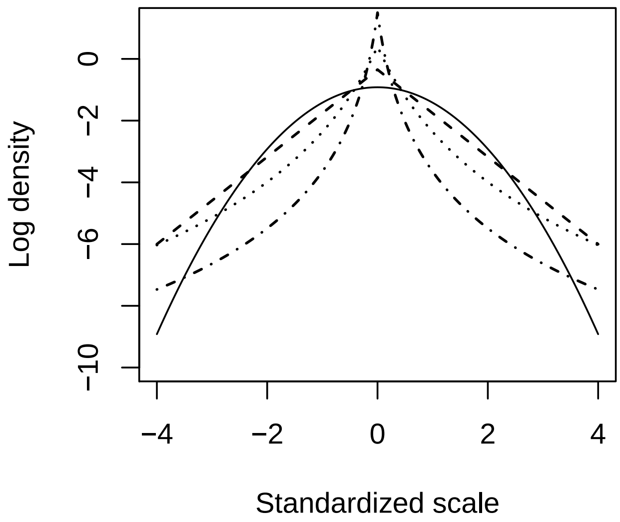

Figure 3: The figure shows the logarithm of the density of a normal distribution (full drawn), the double exponential (dashed) and an instance of the double Pareto distribution (dot-dashed). The figure illustrates that there is a gap in the functional form of the normal distribution on one side and the double exponential distribution and class of double Pareto distributions on the other, that is, there is no member of the latter class that resembles the normal distribution.

impression of how an observed distribution deviates from a specified probability distribution. Due to the nature of some of these distributions, semi-log and log-log plots of tail probabilities are used. For the double Pareto distribution, some scholars make plots of log of the tail probability for binned data against the midpoints of the bins. Then a separate value of the slope β is estimated from a regression analysis for the left and right tails. An example is shown in the right part of Figure 4. As the figure illustrates, the estimation need not be optimal since estimated tail probabilities can be larger than one. Formally, this approach corresponds to the double Pareto (or double Lomax, Lomax (1954)) model

$$P(X \leq x) = \begin{cases} (1-u)\left(1+\frac{|x|}{\delta}\right)^{-\beta_1} & x < 0, \\ 1-u\left(1+\frac{x}{\delta}\right)^{-\beta_2} & x \geq 0, \end{cases} \tag{2}$$

with δ chosen to be half the bin width and with $0 < u < 1$ being the fraction of positive increments. Padgett (1980) was the first to suggest this model in the case $\beta_1 = \beta_2$, and used a value of $u \neq \frac{1}{2}$ to incorporate skewness in the data. However, the price of having $u \neq \frac{1}{2}$ is a jump in the density at zero. This is seen in the left part of Figure 4 where the logarithm of the estimated density corresponding to (2) is shown, both for the suggestion of Padgett (1980) ($\beta_1 = \beta_2$) and for the suggestion of Breunig and Jones (2011) ($\beta_1 \neq \beta_2$). In both cases, maximum likelihood estimates, as suggested in Padgett (1980), are used instead of graphical estimation methods. This jump in the density at zero is inconsistent with what we know about empirical budget change distributions (see also Jensen, Mortensen and Serritzlew 2016)

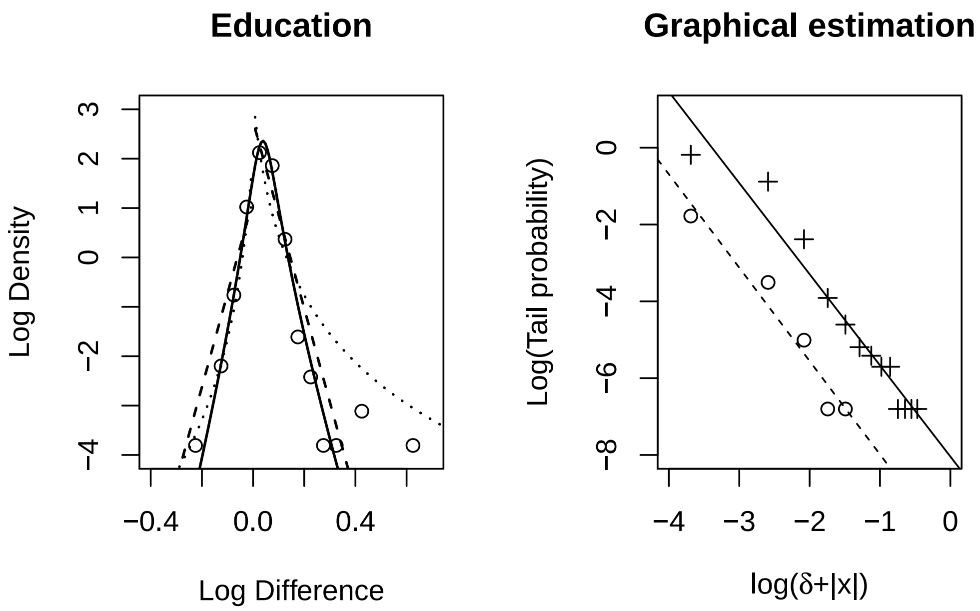

Figure 4: US budget data on Education. The left part shows a log-histogram (circles) together with the estimated double Pareto of Padgett (1980) (dashed) and the double Pareto of Breuning and Jones (2011) (dotted). Both densities have a jump at zero illustrated by the vertical lines to the left of the plot. The right part illustrates a graphical estimation procedure based on linear regression of the logarithm of the tail probability on $\log(\delta + |x|)$, where δ is half the bin width (right tail: plusses, left tail: circles). For the right tail, the estimated line is above zero at the first point: being a log probability, the line ought to be below zero.

Obviously, some simplification is necessary when it comes to comparing and ultimately explaining variation across budget distributions. However, to approximate the diversity of public budget distributions, we need a model with at least four parameters. Two of these parameters represent location and scale, for example through the mean value and the standard deviation. This is sufficient for a gaussian distribution, but more is needed in the nongaussian case. Like in descriptive statistics, where one uses the skewness and kurtosis, we need at least two parameters to describe the form of the distribution. Another way of looking at this is that a parameter is needed for each of the two tails of the distribution. There is no a priori reason to expect symmetry of the distribution (zero skewness), and restricting to such a subclass may hide important aspects of the distribution. In fact, asymmetry is common in budget distributions as well as in many other distributions of policy and administrative data. The comparative distributional method, which is introduced in the next section, builds on such a four-parameter model.

2 The advantages of adopting a CDM approach

The comparative distributional method is a powerful analytical tool that overcomes many of the limitations of existing distributional methods. We find that the CDM is a statistical tool that can closely approximate a broad range of empirical budget change distributions and provide an exact mathematical expression for the corresponding density. In this way the CDM moves the study of budget distributions forward by allowing for inferential statistics, statistical modeling of distributions, and prediction. This is a necessary condition to pave the way for advancing theories about budget change distributions as well as for the use of distributional analysis in public administration more generally.

The proposed method builds on the Normal Inverse Gaussian (NIG) class of distributions. This class of distributions has been used in literatures on phenomena with heavy tailed distributions in different areas such as the study of finance, turbulence, and speech enhancement (Bolviken and Benth 2000, Fragiadakis, Karlis and Meintanis 2009, Kalemanova, Schmid and Werner 2007, Stentoft 2008, Barndorff-Nielsen, Blæsild and Schmiegel 2004, Hendriks and Martin 2007). This means that the NIG class of distributions is well understood, and that powerful statistical tools already exist and are readily available to perform inferential tests of differences between budget distributions, to estimate models, and to make predictions.

We argue that the NIG class of distributions is relevant to the distributional study of public budgeting because of its flexibility and useability. By flexibility we mean that the NIG class of distributions allows for a wide range of distributional forms with a normal distribution at one end, a Cauchy distribution at the other end, and a large range of skewness and kurtosis. Unlike the exponential and the Pareto models applied previously in the public budget literature, the NIG distribution has four parameters that can flexibly represent distributions with fat tails and high skewness. The kurtosis spans the full range from zero (the normal distribution) to infinity (the Cauchy distribution). A particular value of kurtosis and skewness points to a particular NIG distribution (up to scale and location changes), and the form of the NIG distributions empirically fits heavy-tailed budget distributions well. In other words, a wide range of empirical budget distributions can be modelled precisely using the four parameters of the NIG distribution.

By useability we mean that it is easy to estimate the parameters of the NIG distribution. In early studies, scholars constructed their own computer code for finding the maximum likelihood estimates. Examples include studies of stock returns (Bolviken and Benth 2000), changes in exchange rates (Fragiadakis, Karlis and Meintanis 2009), and turbulence research (Barndorff-Nielsen, Blæsild and Schmiegel 2004). Today, one can simply use the estimation algorithm in one of the R-packages *fBasics*((Wuertz et al. 2017)) or *ghyp*. In Supplementary Material SM F, we provide a hands-on guide on how to conduct the type of analyses shown in this paper.

3 The NIG distribution and its properties

In this section, we present the NIG distribution and its basic properties, which constitute the core of our proposed methodology. Readers who are mainly interested in the application of the comparative distributional method may skip this section and go to the following sections.

The normal inverse Gaussian distribution has four parameters, α, β, δ and μ. Here $\delta > 0$ is a scale and μ a location in the sense that if x has the NIG$(\alpha, \beta, \delta, \mu)$ distribution, the transformed variable $a + bx$ is NIG$(\alpha, \beta, b\delta, a + b\mu)$. The parameters α and β are invariant under this transformation and define the *shape* of the distribution. The range is $0 \leq |\beta| \leq \alpha$ and $\alpha + \beta$ and $\alpha - \beta$ determine the left and right tail of the distribution, respectively. Letting $y = (x - \mu)/\delta$ the density of y is

$$h(y; \alpha, \beta) = \frac{\alpha K_1(\alpha\sqrt{1+y^2})}{\pi\sqrt{1+y^2}} e^{\sqrt{\alpha^2-\beta^2}+\beta y}, \tag{3}$$

with K_1 being the modified Bessel function of the second kind.

Barndorff-Nielsen (1997) introduced the distribution and chose the name because when x has the NIG$(\alpha, \beta, \delta, \mu)$ distribution, we can write $x = \mu + (\delta\beta/\gamma)z + (\delta/\sqrt{\gamma})\sqrt{z}u$, where $\gamma = \sqrt{\alpha^2 - \beta^2}$, u has a standard normal distribution, and z has an inverse Gaussian distribution with mean 1 and variance $1/\gamma$. Thus, for a fixed value of z, we have a normal distribution with the mean and variance dependent on z. However, when z varies, we get a mixture of different normal distributions. The degree of mixing can directly be measured through the degree of variation in z. The classical way of measuring the degree of variation is through the *coefficient of variation* (cv) defined as the standard deviation divided by the mean (also known as the relative standard deviation). Intuitively, we can say that the coefficient of variation for z measures the ratio of the variances for x between years with a high variance and years with a low variance. For the mixing variable z, the coefficient of variation is $\mathrm{cv} = 1/\sqrt{\gamma}$. For convenience, we transform the shape parameters α and β to a bounded region, called the *shape triangle*, using

$$\xi = \frac{\mathrm{cv}}{1 + \mathrm{cv}} = \frac{1}{1 + \sqrt{\gamma}} \quad \text{and} \quad \chi = \frac{\beta}{\alpha}\xi,$$

with range $0 \leq |\chi| < \xi \leq 1$ supplemented with the boundary points $(\chi, \xi) = (\pm 1, 1)$ and

$(0,1)$. Figure 5 depicts the NIG density (3) at various positions in the shape triangle, and Figure 6 illustrates how (χ, ξ) relates to skewness and kurtosis.

In the bottom of the shape triangle, $\xi \to 0$, we find the normal distribution. Approaching the right side of the triangle, corresponding to a fixed value of ξ and $\chi \to \xi$ and taking $\mu = 0$ and the variance equal to 1, the limiting density is an inverse gaussian density with mean $\sqrt{1-\xi^2}/\xi$ and unit variance. Approaching the top of the triangle, corresponding to $\alpha \to 0$, we have the Cauchy distribution where tail probabilities decay as $1/x$. In the top corners, we have the cases with $\beta = \alpha$. For a distribution inside the shape triangle, the tail probability decays faster than a power law for sufficiently large x. For a fixed value of ξ, skewness is proportional to χ, and a fixed value of the kurtosis gives a parabolic-shaped region in the shape triangle. Supplementary Material SM C contains further information on the NIG distribution.

4 A comparative distributional analysis of public budget distributions

To demonstrate the comparative insights that can be gained from the CDM approach, we utilize budget distributions covering several countries, several levels of aggregation, as well as both national and subnational political institutions. More specifically, we re-analyze the rich set of public spending data used in Jones et al. (2009), which constitutes the to date most comprehensive analysis of public spending distributions.[3] We use log diffences for all pairs of consecutive years with positive budget values, that is, $d_{t+1} = \log(b_{t+1}) - \log(b_t)$ where b_t is the budget value in year t. The percentage change is found from the log difference as $\exp(d_{t+1}) - 1$. The complete dataset has 47,394 budget entries of which 505 are missing and 27 are nonpositive. Table 1 provides an overview of the data sets.

[3]We are very grateful to the multiple authors of the Jones et al. (2009) article for the generous and unrestricted permission to use their spending data in this paper.

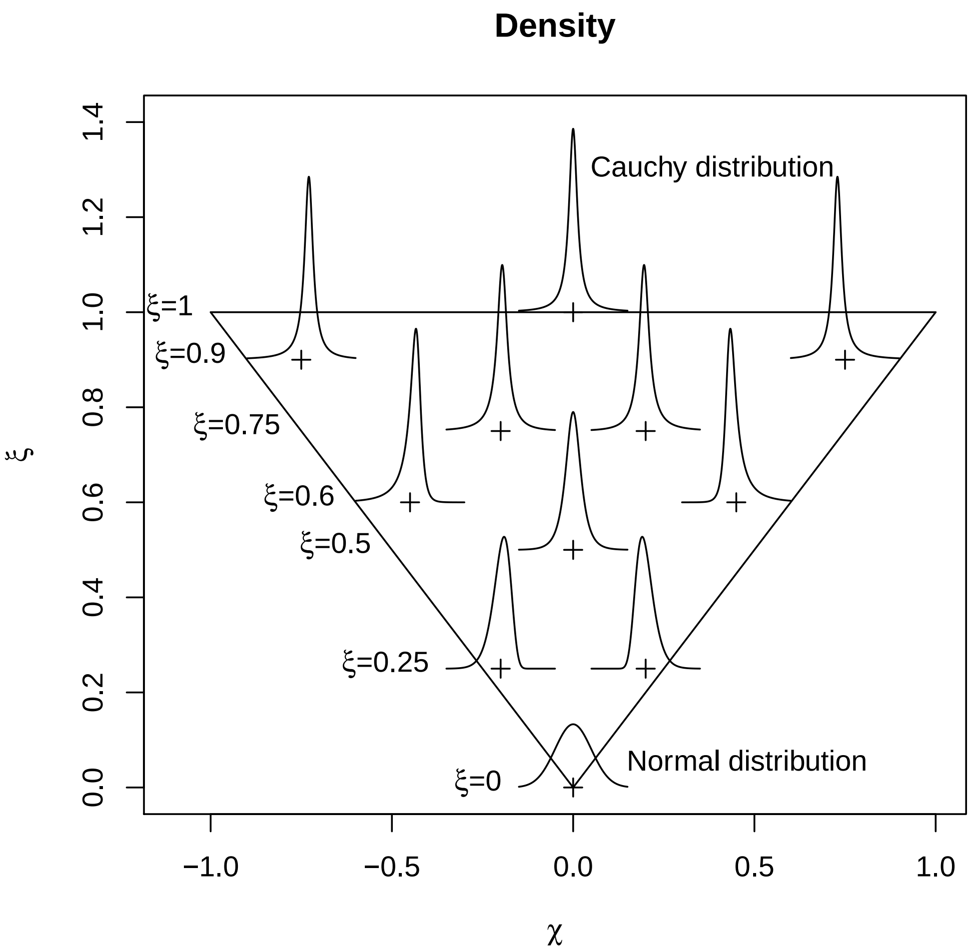

Figure 5: Plot of densities for various values of the shape parameters. The values of ξ are shown in the figure, and the values of χ used are 0, ± 0.2, ± 0.45 and ± 0.75. The distribution is scaled so that the range from the 16th to the 84th percentile is fixed and centered to have mean zero. For each density, the origin (+) of the coordinate system is at (χ, ξ).

We first estimate the distributions for each data set and show how the NIG shape triangle can reveal new important insights that cannot be detected by descriptive statistics alone. We then provide a comparative distributional analysis of whether differences between the distributions are statistically significant. Next, we show how density functions and cumulative distribution functions can be estimated and show how these functions can be used to predict the likelihood of any budget change event. Neither of these calculations are possible without using the NIG-based modeling approach. Finally, we compare

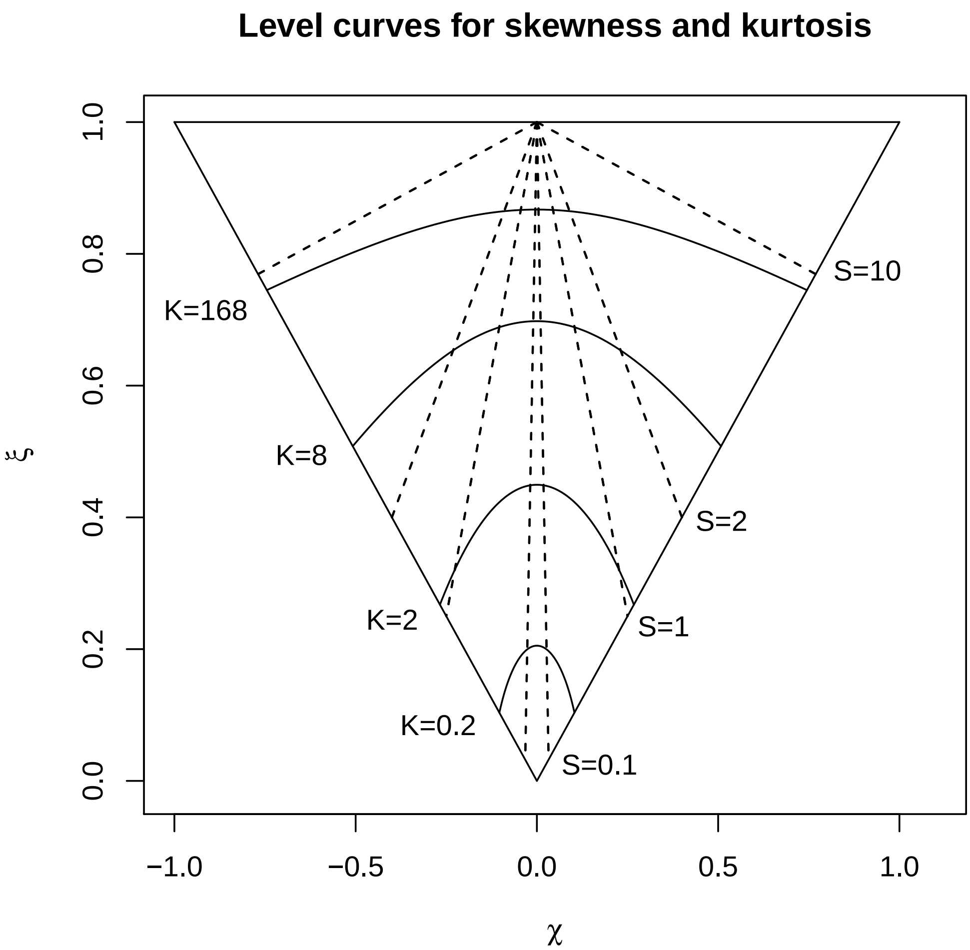

Figure 6: Level curves for skewness (dashed) and kurtosis (solid). The kurtosis values are shown as K, and curves with positive skewness are shown as S.

the performance of the NIG model with two other distributional models that have been employed in the budget literature. We use the chisquare goodness of fit statistic and find that the NIG model consistently fits better. In many of the larger data sets, the NIG model provides a very strong fit, whereas the alternative models give a poor fit.

The left part of Figure 7 shows the shape parameters for the seven national-level data sets in a NIG shape triangle. All data sets have ξ-values above 0.65 indicating a high kurtosis. The χ-values for the distributions differ, however, with a relatively low value for the Canadian data set and a relatively high value for the Belgian data set, pointing to a left skew and a right skew dataset, respectively.

Table 1: Data Descriptions. Abreviations: N: Number of annual change observations; *OMB*: OMB programmatic subfunctions; *F*: functions; *P*: programs; *S*: states; *M*: ministries; *Mu*: municipalities

Dataset	Type	Date	Units Pooled	N
National Governments (long series)				
United States	Outlays	1800-2005		187
US Domestic	Outlays	1800-2004		187
US Defense	Outlays	1800-2004		187
France	Outlays	1820-2002		182
National Governments (pooled)				
United States	Budget Authority	1947-2005	52 OMB	3137
France	Outlays	1868-2004	7 M	1063
Germany	Outlays	1962-2000	26 F	982
Great Britain	Outlays	1981-1999	14 F	252
Belgium	Outlays	1991-2000	27 F	243
Denmark	Outlays	1971-2003	26 F	832
Canada	Outlays	1990-2004	12 F	144
Subnational Governments (pooled)				
U.S State	Operating Outlays, 10 F	1984-2002	50 S	9000
Danish Local	Operating Outlays, 4 F	1991-2005	265 Mu	15174
Programmatic spending series				
United States	Budget Authority, 52 OMB	1947-2005		43-56
US State	Operating Outlays, 10 P	1984-2002	50 S	900
Danish Local	Operating Outlays, 4 P	1991-2005	Mu	3794

Jones et al. (2009) analyzed these countries as well and found a positive correlation between a measure of institutional friction (a summary measure of institutional measures of e.g. executive dominance) and L-kurtosis. We can reproduce these results, with a similar relationship between the same measure of friction as used in Jones et al. (2009) and the shape parameter ξ, see Table SM 1 in the Supplementary Material.

Interestingly, the plot shows that friction may also be related to the skewness parameter χ in the sense that countries with high friction also tend to have higher χ values than coutries with lower friction. Given that a low χ means that the distribution is skewed to the left, these results suggest that large cutbacks may be more likely in countries with high friction than in countries with lower friction. Conversely, large increases may be more likely in countries with higher friction than in countries with low friction (see also Table SM 1 in the Supplementary Material). We see no reason why a distributional analysis should be limited to focusing on one of the moments, and as discussed in the concluding part of the paper, the integration of skewness and kurtosis into one analysis can lead to new theoretical insights.

The right part of Figure 7 shows the shape triangle for the 10 spending items in US states. The ξ-values vary more than in the aggregated national data sets in the top of the figure. Spending items such as Education and Highways have lower ξ-values than Parks and Hospitals, indicating lower levels of kurtosis. One item, Parks, has a quite low χ (skewed to the left) while the nine other items have more symmetric distributions.

While Figure 7 shows that there is considerable variation in the estimated ξ- and χ-values, particularly for the ten US spending items in the right part of the figure, it is not possible to judge from the figure whether the difference between the estimates is statistically significant. However, each point does have an oval-shaped confidence interval around it making it possible to evaluate statistical significance. In Table 2 we have performed a test of the hypothesis that the shape parameters (χ, ξ) for dataset 1 equal the shape parameters for dataset 2, and the table entry is the p-value of the test based on the asymptotic χ^2-distribution of the log likelihood ratio statistic. We use the standard 0.05-limit for statistical significance.

As an example, the diffence in shape parameters for Highways and Natural Resources is statistically significant with $p < 0.03$. In fact, the shape parameters for Highways are significantly differently located than for all other spending items, except Education. Even the location of Police, which is central among the points in the shape triangle, is statistically different from almost half of the other spending items. Note that these significance tests represent a crucial contribution to previous comparative budget research, which either lacked inferential techniques or has been restricted to categorical comparisons between the normal, the exponential, or the paretian distribution.

Referring back to the message in Figure 2, we can illustrate the relevance of the kurtsosis measure critique with an empirical example. *Hospitals* and the spending category *Others* have the same shape (p-value is 0.20), but the two kurtosis values are 172 and 8.3, respectively. Similarly, *Education* and *Corrections* have statistically significant different shapes although the two kurtosis values are similar: 17.9 and 16.2, respectively.

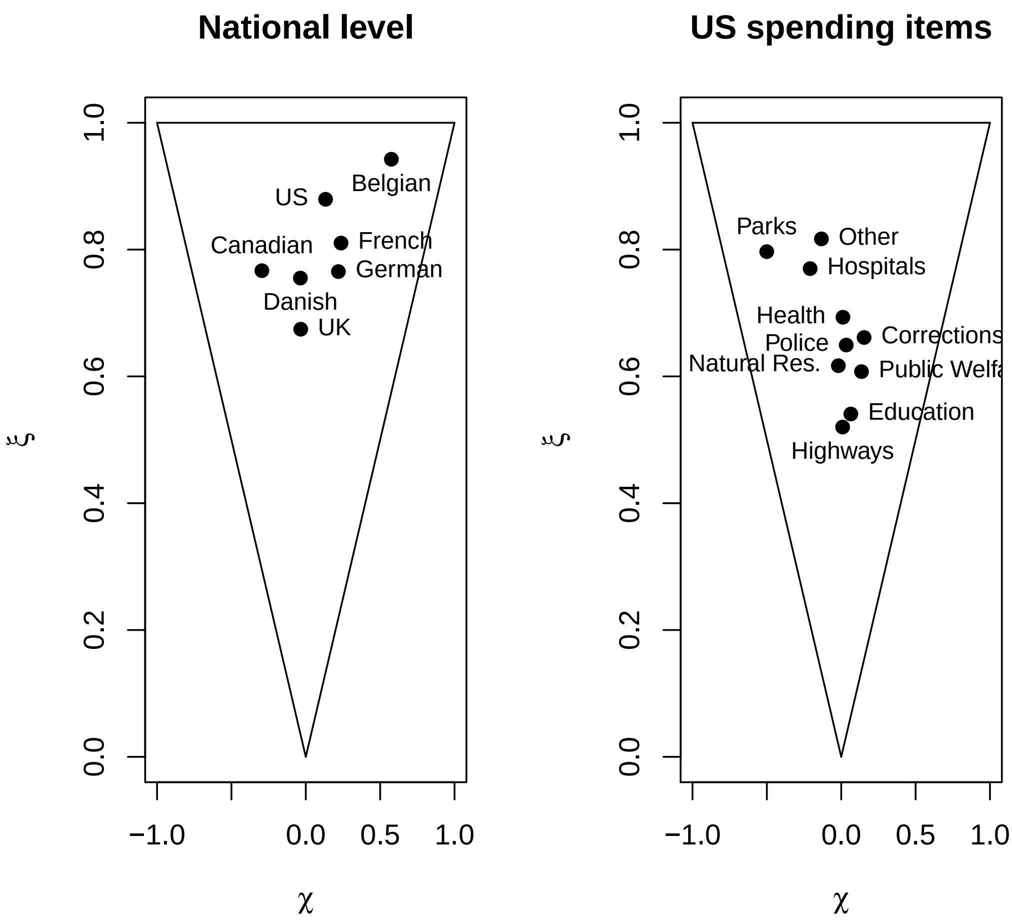

Figure 7: Shape triangles for public spending data sets

We now turn to examples of the additional insights gained by estimating specific NIG distributions. Figure 8 displays histogram and log histogram for Highways and Hospitals in US Congressional budget authority from 1947-2012. For Hospitals, the histogram shows the familiar pattern with a high peak, slender shoulders, and fat tails, which is less pronounced for Highways. The descriptive kurtosis measures confirm that these data follow a leptokurtic distribution (L-kurtosis is 0.22/0.43 (Highways/Hospitals) and kurtosis is 2.47/172. [4] The estimated NIG distribution is overlayed the observed distribution.

[4]Note, as we explained above, that we use excess kurtosis, which is the fourth central moment divided by the square of the variance, subtracted the number three. It is also common to report kurtosis values

Table 2: P-values for testing same shape (same value of (χ, ξ)) of two spending item distributions. Column headings are abbreviations of the row names.

Item	PW	Hi	NR	He	Ho	Po	Co	Pa	Other
Education	0.09	0.49	0.03	0.00	0.00	0.00	0.00	0.00	0.00
Public welfare	–	0.01	0.05	0.00	0.00	0.10	0.24	0.00	0.00
Highways	–	–	0.03	0.00	0.00	0.00	0.00	0.00	0.00
Natural Resources	–	–	–	0.05	0.00	0.46	0.03	0.00	0.00
Health	–	–	–	–	0.01	0.34	0.10	0.00	0.00
Hospitals	–	–	–	–	–	0.00	0.00	0.04	0.20
Police	–	–	–	–	–	–	0.27	0.00	0.00
Corrections	–	–	–	–	–	–	–	0.00	0.00
Parks	–	–	–	–	–	–	–	–	0.01

Based on visual inspection, the NIG distribution follows the observed distribution closely.

The estimated NIG distribution gives us the foundation for prediction. It is possible to estimate both the density function and the cumulative distribution function. These functions, estimated for Highways and Hospitals, are seen in Figure 9. The estimation of the NIG distribution thereby allows us to provide an exact mathematical expression for the density. For the case of Highways, this expression is:

$$f(x) = \frac{18.20 \cdot K_1\left(0.851 \cdot \sqrt{1+z^2}\right)}{\sqrt{1+z^2}} \cdot \mathrm{e}^{0.0152 \cdot z}, \quad z = \frac{x - 0.03024}{0.1095}.$$

This has not been seen before in the study of public budgeting. The mathematical expression provides a much more precise characterization of the empirical distribution than summary measures such as kurtosis or L-kurtosis. The expression produces one and only one distribution, and the expression makes it possible to estimate the predicted occurence of any range of annual percentage changes. For instance, the estimated distributions imply that Highways have 10.7 percent of the budget growth rates between 10 and 15 percent compared to 6.7 percent for Hospitals. Public Welfare and Education are the items with lowest likelihood of dramatic reductions (the chance of a reduction of 10 percent or more is 0.026 and 0.008, respectively). For comparison, the chance of a similar cutdown is

without subtracting three; in this case, kurtosis would be reported as 5.47, indicating a weakly leptokurtic distribution.

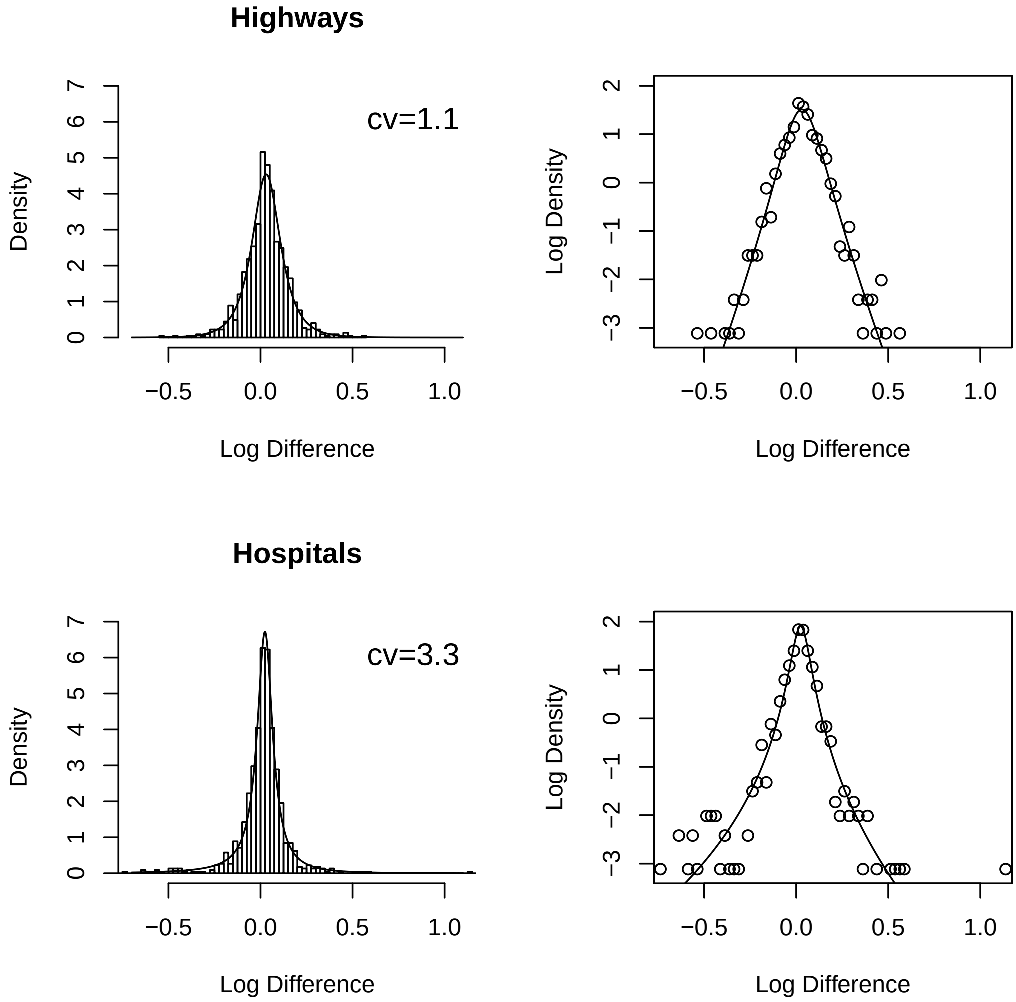

Figure 8: Histogram and log histogram for the US data on Highways and Hospitals with the estimated NIG distribution overlayed. The cv value is coefficient of variation for the mixing distribution. For Hospitals, the ratio of typical high-variance situations to low-variance situations is three times that for Highways.

0.222 and 0.148 for Parks and Others.

This illustrates the detailed insigt in the distributions that can be gained by estimating them. Substantially, the differences across spending areas may be compatible with classic ideas in public administration stressing the different distribution of vested interests across issue areas (Wilson 1980, Mortensen 2005). What the NIG-based approach offers are new opportunities to integrate these ideas into the study of budget distributions.

Finally, we compare goodness-of-fit scores across three different models. For each dataset listed in Table 1, we calculate goodness-of-fit test scores for the NIG-based model,

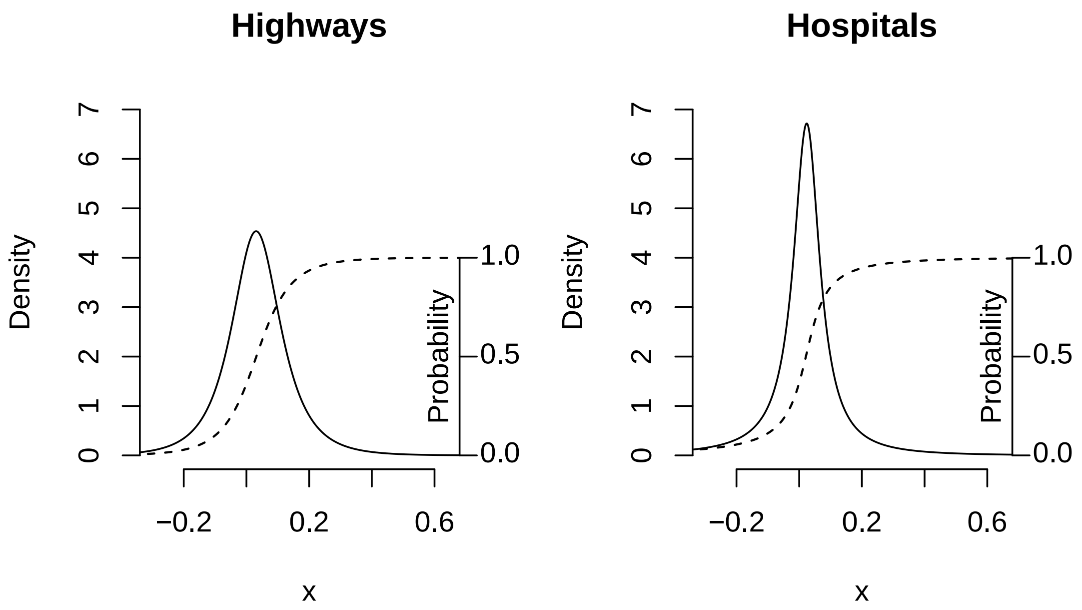

Figure 9: Density and cumulative distribution function for the fitted NIG distributions from Figure 8.

the three-parameter Pareto model used by Padgett (1980) (Pareto model 1), as well as the two-parameter model described by Breunig and Jones (2011) (Pareto model 2).[5]

Table 3 provides a summary across 79 different goodness of fit tests for the NIG-based model compared with the two other models. To perform a goodness of fit test for a particular model, we have used a bin width of 0.05. In order to have expected counts larger than 5, bins are grouped together starting from the extreme tails. The test statistic is $G = 2\sum_i a_i \log(a_i/e_i)$, where a_i is the observed count, and e_i is the expected count. Table entries are $G/(K-1-d)$, where K is the number of counts, and d is the number of parameters of the model.[6] The division by $K-1-d$ ensures that a model with two parameters that fits the data does not generally have a larger value of the test statistic as compared to an extended model with more parameters. Lower values of the test statistic indicate a better fit of the distribution to data.

[5]The two-parameter Pareto model has two parameters corresponding to the slopes of the two tails, but actually needs a third parameter to put together the left and right tails.

[6]Note that in the construction of the test statistic G, the parameter values that minimize G are used rather than the maximum likelihood estimates from the original log differences. However, when we report parameter values and plot these in the shape triangle, the latter are used.

Several conclusions can be drawn based on Table 3. Across all types of budget distributions, the NIG model performs better than the Pareto models when we consider the average of the goodness of fit statistics. The same holds true when we consider the spread, the smallest value, and the largest value of the test statistics. When we look at the individual values behind the summary values in Table 3, we see that in 24 out of 27 datasets with more than 100 observations the NIG model has a better fit score, and in many of these cases, the NIG-based model provides a fine fit, whereas the Pareto models provide a poor fit. For the 52 dataseries with less than 100 observations, both the NIG model and the first Pareto model generally fit the data. The NIG model has the lowest score in 29 of the 52 cases, and even in the few cases where Padgett's (1980) version of the Pareto model has the lowest test score, the NIG model obtains low test scores.

The latter points to the standard deviations of the test scores. Here it is worth noting the very low standard deviations of the NIG models across all types of spending data distributions. These low scores corroborate the argument that the class of NIG distributions is *flexible* in the sense that it can fit a broad range of empirical distributions. In other words, the low variability in test scores shows that the NIG model can almost always provide a good approximation of the empirically observed budget distributions, whereas the Pareto models sometimes are widely off target.

In the test statistic, the different number of parameters in the models has been taken into account, and the better fit of the NIG-based model in Table 3 therefore reflects that we do indeed need all four parameters of the NIG model to describe data. As argued above, four parameters are needed to have a general model with a position and a scale parameter and two parameters to model skewness and kurtosis. Having less parameters potentially hides important aspects of data.[7]

[7]This is equivalent to the familiar case where a one-parameter model $y = bx$ can be highly misleading if the true model is $y = a+bx$.

Table 3: Goodness-of-fit tests

		Test scores			
Type of data sets	Number of data sets	Mean	Sd.dev.	Min	Max
National Governments (long series)					
NIG	4	1.04	0.92	0.33	2.28
Pareto model 1	4	1.34	0.95	0.70	2.71
Pareto model 2	4	3.32	1.65	2.11	5.64
National Governments (pooled)					
NIG	7	1.10	0.47	0.27	1.64
Pareto model 1	7	1.24	0.52	0.41	1.97
Pareto model 2	7	5.08	2.98	1.49	10.09
Subnational Governments (pooled)					
NIG	2	5.13	2.32	3.49	6.78
Pareto model 1	2	10.5	4.00	7.67	13.33
Pareto model 2	2	51.2	4.27	48.18	54.22
Programmatic spending series					
NIG	52	1.25	0.74	0.07	3.50
Pareto model 1	52	1.71	1.21	0.13	5.91
Pareto model 2	52	3.67	3.16	0.29	17.93
10 US State level programs pooled across years and states					
NIG	10	1.63	0.93	0.36	3.76
Pareto model 1	10	7.29	9.16	1.77	31.86
Pareto model 2	10	28.66	21.85	10.04	81.68
4 Danish Local level programs pooled across years and municipalities					
NIG	4	1.78	0.79	0.88	2.77
Pareto model 1	4	8.11	12.13	0.97	26.17
Pareto model 2	4	38.65	43.88	1.07	101.75

5 Conclusion and discussion

Distributional analysis was introduced to public administration in the study of public budgets. Before turning to the question of the potential of the CDM in other fields within public administration, it is useful to summarize its contributions and usefulness for public budgeting. The starting point of this paper was that the distributional turn in public budgeting has transformed the approach to public budgeting, but in order to fulfill its potential, scholars need the statistical tools to perform comparative distributional analysis. With the toolbox of distributional analysis used previously, we can reject the classic incrementalism model of normally distributed year-to-year changes. We also know that Padgett's (1980) serial judgement model as well as Jones and Baumgartner's (2005b) dynamic model of choice for public policy are more consistent with observed output distributions than the incrementalism model (Jensen, Mortensen and Serritzlew 2016). Yet, a broad set of alternative models may be just as or even more consistent with observed budget distributions.

Thus, further theoretical development is definitely warranted, but without empirical tools to adequately examine such theoretical models, they will likely remain mere speculation. The standard methods only allow for quite rough empirical testing, and they are incapable of clearly distinguishing between distributions of different nature. This restricts the basis for advancing theoretical models. Statistical models must be sufficiently flexible to account for differences in data implied by theoretical models. To put it bluntly, if a hammer is your only tool, there is a risk that everything will look like a nail. Precise empirical description is a prerequisite for new theoretical insights. Therefore, it is crucial to be better able to characterize empirical budget distributions, and the CDM is a strong tool for this.

As emphasized in this paper, the CDM allows for inferential statistics, statistical modeling of distributions, and prediction. Inferential statistics makes it possible to compare budget distributions from different political systems and different policy areas to see if they are statistically separable. With this statistical model it is also possible to estimate parameter values that precisely characterize a distribution, and to predict the likelihood of specific changes in budgets. It has also been shown in this paper how the underlying NIG-based model fits a large number of empirical budget change distributions from a wide selection of countries and from various tiers of government.

As illustrated by the NIG shape triangles, both skewness and kurtosis can easily be integrated into the same distributional analysis, which opens up new possibilities both for theorizing and testing arguments about kurtosis and skewness. This is an important addition to existing tools that can lead to a better integration of the various research traditions within public budgeting. Whereas much of this literature has been concerned with the direction of budgetary changes, often arguing for a systematic pressure towards increased public spending, another strand of the literature has mostly focused on the effect of cognitive constraints and disproportional information processing. Most likely both aspects are important, but whereas the latter has led to a focus on the kurtosis of distributions, the former perspective has clear implications for the mean and skewness of the distributions.

Given that the mixture idea is fundamental to the NIG approach, the better fit of the NIG models also implies that scholars may develop better theoretical models by further scrutinizing the mixture aspects of public budgeting. Public budgets are the result of complex underlying decision-making processes ranging from (almost) automatic mandatory spending to highly discretionary spending with little oversight and control. This mixture of decision-making processes is fundamental to public budgets and should therefore play a central role in both theorizing and in the empirical tools used to characterize and analyze budget distributions.

In our view, the CDM also has the potential to advance our understanding of a range of other public administration topics besides public budgeting. One example is the big data movement in public administration, which is made possible by better access to large data sets, particularly from register data available from databases and from textual output available from social media and websites. Such large data sets have been utilized, for instance, in research on dynamics of laws and executive orders, which is typically based on thousands of rule-change observations (Jakobsen and Mortensen 2015, March, Schulz and Zhou 2000, Schulz 1998, van Witteloostuijn and Jong 2007, Zhou 1993). This research has focused on average changes, but new theoretical questions, and new empirical answers can be gained from systematic comparisons of distributions of rule changes across different types of rules, across different rule-producing organizations, different policy areas, and perhaps different countries.

Relatedly, earlier research has shown that change distributions from congressional hearings, law-making activity, parliamentary debates, etc. are shaped rather similarly to budget distributions (Baumgartner et al. 2009). Furthermore, a new and promising research agenda has started to compare decision-making outputs across democratic and authoritarian regimes as well as across public and private organizations (Epp 2018). Based on the comparative distributional tools advocated in this paper, it is possible to better identify and characterize the different distributions in order to improve our understanding of how the decision-making processes of public budgets deviate from the processes behind hearing activities, elections, and media, and more generally, how policymaking differs across democratic and non-democratic regimes.

The study of public sector reform is another research field that may benefit from distributional analysis (see e.g. Boyne, Farell and Law 2003, Jakobsen 2013). We consider examples of public school reform, incentive systems, and co-production. Lavy (2016), for instance, examines the effects of performance-for-pay on student attainment. To understand how student performance is affected by a particular type of school reform, it can be useful to have a precise estimation of the full distribution of student performance, not just the average effects. Register data on individual-level student test scores in primary schools or performance data for public employees cannot be completely summarized by descriptive statistics. To understand how different groups respond to, for instance, the introduction of performance-for-pay, school vouchers, or school-level incentives systems, it is relevant to study distributions of student performance. This allows us to answer questions about, for instance, the effects for the number of top students and for students in risk of failing, and how this differs among groups, rather than just test-score averages. To estimate the effects of reform on such distributions, the comparative distributional method is required.

Similarly, to understand how incentive systems affect employee performance, it is useful to study the distribution of performance data. The CDM will allow for an understanding of whether an incentive system has a statistically signicant effect on the distribution of performance, including exact estimates of changes in the number of top-performing employees.

As a final example of public sector reform research that may benefit from a CDM approach, co-production has been promoted because involving citizens as producers can lead to considerable gains in public service provision efficiency when input from the public sector and from citizens is complementary (Parks et al. 1981, Ostrom 1996). Studies have shown that these gains in efficiency can be considerable. Jakobsen and Andersen (2013) find that involving parents in educational development has a positive impact on educational outcomes. However, the effects are not uniformly distributed; the co-production initiative has stronger effects on disadvantaged families (Jakobsen and Andersen 2013, 709). Hence, not only average effects but also the distribution of these effects are important. The CDM can be used as a tool to understand how different types of co-production initiatives can lead to different distributions of outcomes, giving insights about the effects for the top-performing students, and students performing below important thresholds.

We believe that these examples also illustrate a more general application of distributional analyses. Distributions of outcomes such as student performance improve our understanding of the differential effects of government initiatives. This is important. Public sector reforms should not only be evaluated based on their average effect on performance. Equity - or equality - is another crucial aspect of performance (Boyne, Farell and Law 2003). More focus on distributional effects, which can be studied by the CDM, will allow for better insights into this aspect of public sector performance.

References

Andersen, Simon Calmar and Peter B. Mortensen. 2010. Policy Stability and Organizational Performance: Is there a relationship? *Journal of Public Administration Research and Theory* 20 (1): 1-22.

Barndorff-Nielsen, O.E. 1997. Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling. *Scandinavian Journal of Statistics* 24:1-13.

Barndorff-Nielsen, O.E., P. Blæsild and J. Schmiegel. 2004. A parsimonious and universal description of turbulent velocity increments. *The European Physical Journal B* 41:345-363.

Baumgartner, Frank R., Christian Breunig, Christoffer Green-Pedersen, Bryan D. Jones, Peter B. Mortensen, Michiel Nuytemans and Stefaan Walgrave. 2009. Punctuated Equilibrium in Comparative Perspective. *American Journal of Political Science* 53 (3): 603-620.

Bendor, Jonathan. 2003. Herbert A. Simon: Political Scientist. *Annual Review of Political Science* 6: 433-471.

Bolviken, E., and Fred Espen Benth. 2000. "Quantification of risk in Norwegian stocks via the normal inverse Gaussian distribution." *Proceedings of the AFIR 2000 Colloquium, Tromso, Norway*: 87-98.

Boyne, G.A., Catherine Farell and Jennifer Law. 2003. *Evaluating public management reforms: Principles and practice*. McGraw-Hill education.

Breunig, Christian, Chris Koski and Peter B. Mortensen. 2010. Stability and Punctuations in Public Spending: A Comparative Study of Budget Functions. *Journal of Public Administration Research and Theory* 20 (3): 703.

Breunig, C. and B.D. Jones. 2011. Stochastic Process Methods with an Application to Budgetary Data. *Political Analysis* 19 (1): 103-117.

Davis, O.A., M.A.H. Dempster and A. Wildavsky. 1966. A Theory of the Budget Process. *American Political Science Review* 60, 529-47.

Davis, Otto A., M.A.H. Dempster and Aaron Wildavsky. 1974. Towards a Predictive Theory of Government Expenditure: US Domestic Appropriations. *British Journal of Political Science*, 419-52.

Epp, D. 2018. *The Structure of Policy Change*. Chicago: University of Chicago Press.

Flink, Carla M. 2018. Predicting Budgetary Change: The Effect of Performance Gaps. *Journal of Public Administration Research and Theory* Advance Access.

Fragiadakis, K., D. Karlis and S.G. Meintanis. 2009. Tests of fit for normal inverse Gaussian distributions. *Statistical Methodology* 6, 553-564.

Hendriks, Richard C. and Rainer Martin. 2007. MAP Estimators for Speech Enhancement Under Normal and Rayleigh Inverse Gaussian Distributions. *IEEE Transactions on Audio, Speech and Language Processing* 15 (3) 918-927

Hosking, J.R.M. 1990. L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal ofthe Royal Statistical Society*, Ser. B, 52: 105-124.

Jakobsen, Morten. 2013. Can Government Initiative Increase Citizen Coproduction? Results of a Randomized Field Experiment. *Journal of Public Administration Research and Theory* 23: 27-54.

Jakobsen, Morten and Simon Calmar Andersen. 2013. Coproduction and Equity in Public Service Delivery. *Public Administration Review* 73 (5): 704-713.

Jakobsen, M.L.F. and Peter B. Mortensen. 1980. How Politics Shapes the Growth of Rules. *Governance* 28, 497-515.

Jensen, Jens Ledet, Peter B. Mortensen and Søren Serritzlew. 2016. The Dynamic model of choice for public policy reconsidered: A formal analysis with an apllication to U.S. budget data. *Journal of Public Administration Research and Theory* 26 (2): 226-238.

Jones, Bryan D., Tracy Sulkin and Heather A. Larsen. 2003. Policy Punctuations in American Political Institutions. *American Political Science Review* 97 (1): 151-169.

Jones, B.D. and F.R. Baumgartner. 2005a. A model of choice for public policy. *Journal of Public Administration Research and Theory* 15 (3) 325-351.

Jones, B.D., and Frank R. Baumgartner. 2005b. *The Politics of Attention: How Government Prioritizes Problems*. Chicago: University of Chicago Press.

Jones, B.D., F.R. Baumgartner, C. Breunig, C. Wlezien, S. Soroka, M. Foucault, A. Francois, C. Green-Pedersen, C. Koski, P. John, P.B. Mortensen, F. Varone and S. Walgrave. 2009. A General Empirical Law of Public Budgets: A Comparative Analysis. *American Journal of Political Science* 53 (4): 855-873.

Jones, Bryan D., L. Zalanyi and P. Erdi. 2014. An Integrated Theory of Budgetary Politics and Some Empirical Tests: The U.S. National Budget, 1791-2010. *American Journal of Political Science* 58 (3): 561-578.

Kalemanova, Anna, B. Schmid and R. Werner. 2007. The Normal Inverse Gaussian Distribution for Synthetic CDO Pricing. *The Journal of Derivatives*, 14, pp. 80-94.

Lavy, Victor. 2016. Teachers' Pay for Performance in the Long-Run: Effects on Students' Educational and Labor Market Outcomes in Adulthood. *NBER Working Paper, No. 20983*.

Lomax, K.S. 1954. Business Failures: Another Example of the Analysis of Failure Data. *Journal of the American Statistical Association*, 49, 847-852.

March, J.G., Martin Schulz and Xueguang Zhou. 2000. *The Dynamics of Rules: Change in Written Organizational Codes*. Stanford, CA: Stanford University Press.

Mortensen, Peter B. 2005. Policy punctuations in Danish local budgeting. *Public Administration* 83 (4): 931-950.

O'Toole, Laurence J. and Kenneth J. Meier. 2003. Plus ça Change: Public Management, Personnel Stability, and Organizational Performance *Journal of Public Administration Research and Theory* 13 (1): 43-64.

Ostrom, Elinor. 1996. Crossing the great divide: Coproduction, synergy, and development *World Development* 24: 1073-1087.

Padgett, J.F. 1980. Bounded Rationality in Budgetary Research. *American Political Science Review* 74, 354-72.

Parks, Roger B., Paula C. Baker, Larry L. Kiser, Ronald J. Oakerson, Ellinor Ostrom, Vincent Ostrom, Stephen L. Percy, Martha B. Vandivort, Gordon P. Whitaker and Rick K. Wilson. 1981. Consumers as coproducers of public services: Some economic and institutional considerations. *Policy Studies Journal* 9, 1001-1011.

Robinson, Scott E., Floun'say Caver, Kenneth J. Meier and Laurence J. O'Toole. 2007. Explaining policy punctuations: Bureaucratization and budget change. *American Journal of Political Science* 51 (1): 140-150.

Robinson, Scott E., Carla M. Flink and Chad M. King. 2014. Organizational History and Budgetary Punctuation. *ournal of Public Administration Research and Theory* 24 (2): 459-471.

Schulz, M. 1998. Limits to Bureaucratic Growth: The Density Dependence of Organizational Rule Births. *Administrative Science Quarterly* 43, 845-876.

Stentoft, Lars. 2008. American Option Pricing Using GARCH Models and the Normal Inverse Gaussian Distribution. *Journal of Financial Econometrics* 6 (4) 540-582.

van Witteloostuijn, A. and Gjalt de Jong. 2007. The Evolution of Higher Education Rules: Evidence for an Ecology of Law. *International Review of Administrative Sciences* 73, 235-255.

Wilson, James Q. 1980. The Politics of Regulation. In *The Politics of Regulation*, James Q. Wilson (ed.). New York: Basic Books, pp. 357-94.

Wuertz, Diethelm, Tobias Setz Yohan Chalabi and Martin Maechler. 2014. fBasics: Rmetrics - Markets and Basic Statistics. R package version 3042.89

Zhou, X. 1993. The Dynamics of Organizational Rules. *Maerican Journal of Sociology* 98, 1134-1166.