Undergraduate Economic Review

Volume 10 | Issue 1 Article 5

2013

Application of SGT Family Distributions in Quasi Maximum Likelihood Estimation

Samuel Dodini
Brigham Young University - Utah, sdodini@gmail.com

Application of SGT Family Distributions in Quasi Maximum Likelihood Estimation

Abstract

In the classical normal linear regression model, ordinary least squares estimators (OLS) will be consistent and achieve the Cramer-Rao lower bound for any unbiased estimators. This paper examines the impact of several other error distributions on the properties of the OLS estimators. Several different types of example data commonly available to students and researchers in economics are used to illustrate the impact of nonnormality, because, in application, the assumption of normality may not hold in empirical testing. Using maximum likelihood, I demonstrate that flexible probability density functions better model the residual distribution of different types of data, which suggests improvements in estimation accuracy. I find that this suggested increase of fit applies to almost all data types, with the scale of these likelihood improvements contingent upon data characteristics specific to individual data sets. I conclude that consideration of these distributions is essential for truly rigorous analysis and that parsimony applies when differences between estimators are not significant.

Keywords

flexible distributions, normality, estimation, Gauss-Markov, regression, central limit theorem, applied econometrics, econometrics, policy research, linear estimators

Cover Page Footnote

A special thanks to Dr. James McDonald for helpful comments and MATLAB programming direction.

Standard ordinary least squares (OLS) estimators in a linear regression framework minimize the sum of squared errors. These estimators will be the Best Linear Unbiased Estimators (BLUE) if the Gauss-Markov assumptions hold and will have the minimum variance of all unbiased estimators if the errors are normally distributed. In practice, many of these assumptions are violated. Heteroskedasticity is common in many cross-sectional data sets, as well as some sort of autocorrelation in time series data. While there are several methods of addressing the violation of these Gauss-Markov assumptions, such as generalized least squares, there are fewer rules of thumb to address non-normality in the residuals, which impacts the efficiency of OLS estimators. This can be especially important in areas of public policy in which billions of dollars depend on the choice of estimator. In essence, I ask the question, “What if there is a *better* estimator?” I compare the efficiency of OLS estimators to maximum likelihood estimators assuming the following error distributions:

1) Student's t Distribution (t)
2) Generalized Error Distribution (GED)
3) Inverse Hyperbolic Sine (IHS)
4) Generalized t (GT)
5) Skewed Generalized t (SGT)

These estimators are often called quasi maximum likelihood or partially adaptive estimators as the regression parameters are estimated along with those of the approximating error distribution. These distributions can be related using the SGT tree relationship in Appendix A.

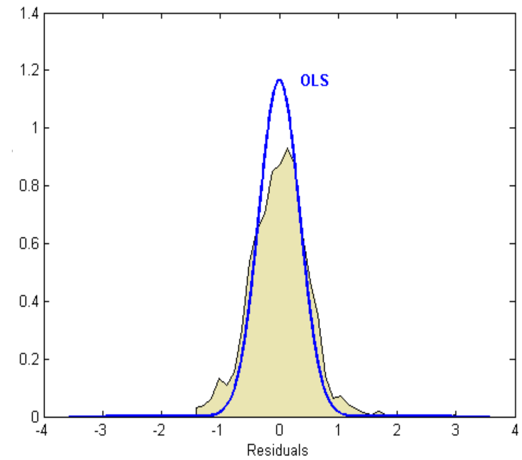
Data

To demonstrate the difference between these several error distributions and the comparative accuracy of their outcomes using quasi maximum likelihood estimation, I examine six separate data sets from the Wooldridge data set collection. These were chosen for their variety, reliable formatting, and workability and provide a diverse framework of possibilities for real world data examination. Summary statistics are provided in Appendix B. Each data set is homoskedastic with no autocorrelation, which isolates the error distribution as a varying factor. I first perform an OLS regression for my dependent variable (the first variable in each data set regressed on all the others) as an initial point of reference.

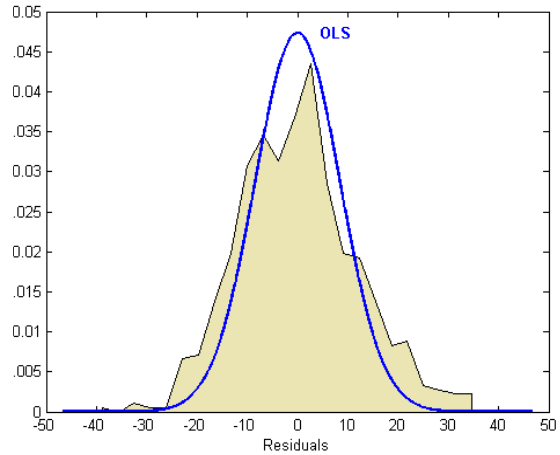
Poorly Matching Residuals

Below are the reported OLS residual graphs for each such regression. These consist of a smoothed histogram of the OLS residuals with an overlaid fitted normal distribution for reference. Notice the discrepancy between these assumed errors and the actual data residuals.

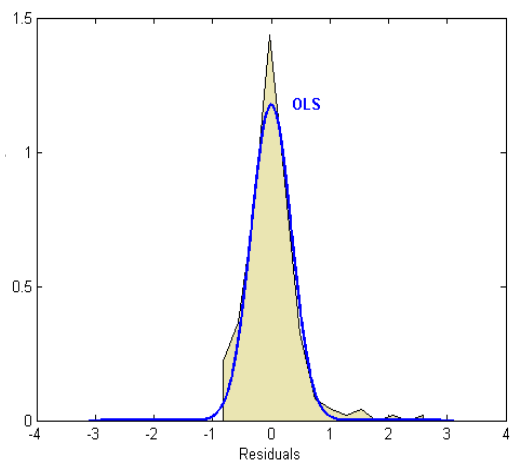
Beauty



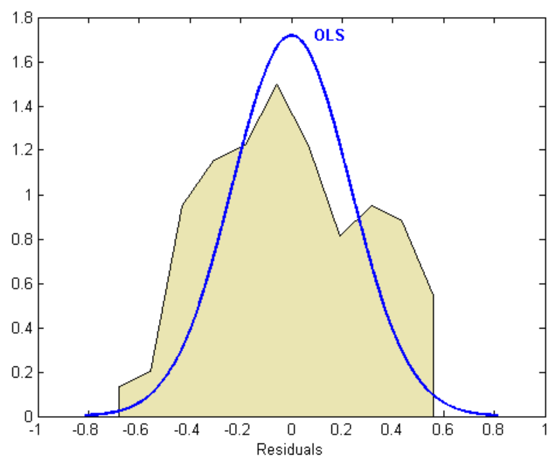
Math Proficiency



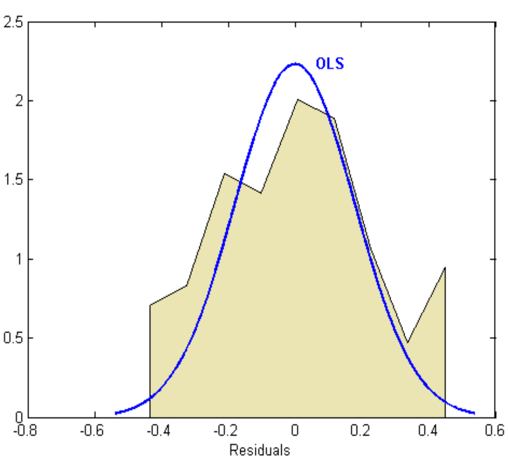
CEO Salary



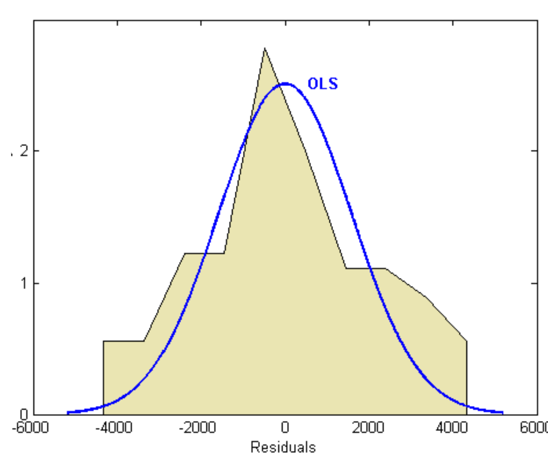
GPA



Crime Rates



Traffic Accidents



The normal distribution does not approximate the actual residuals well due to rigidity issues in skewness or kurtosis, of which kurtosis seems to be the more egregious of the two. In these snapshots, there does not appear to be a specific pattern to these kurtosis issues from these six data sets.

The Current Literature

The essential theme of standard OLS regression theory suggests that, by the Central Limit Theorem, errors *should* be asymptotically normal, which may not be accurate in some specifications. Efromovich (2005) suggests a theoretical justification for the common fall-back of considering residuals as *proxies* for underlying regression errors. However, increased efficiency in computing and econometrics merits delving further into the true errors. Perhaps one of the first papers to examine non-normality in errors in linear regressions was Zeckhauser and Thompson (1970), which examined maximum likelihood estimates using the three parameter power distribution made popular by Box and Tiao (1964). They argue that the "Supposition [of normality] is often unwarranted and... significant gains in likelihood may be achieved when the regression technique allows for the more general class of error distributions," (Zeckhauser and Thompson, 1970). They attribute the inapplicability of the Central Limit Theorem to small sample size, non-normally distributed independent variables, and the presence of the non-random effects of human behavior. They also argue that using variance as a measure of efficiency loses its explanatory power when underlying errors diverge from normality, and is particularly important when facing error distributions with thicker tails. All these make partially adaptive estimation more appealing. Considering even more flexible error terms than Zeckhauser and Thompson sheds light onto how to further tighten estimation in an age of big data and even bigger policy decisions. Higher order moments may be the cause of the discrepancy between what would be predicted by OLS estimation under normality assumptions and residuals with additional parameters.

Perhaps the most compelling evidence for the need for flexible distributions comes from McDonald and Newey (1988), who introduced the Generalized T distribution as a means of tightening estimation for financial data. As my results will show, this distribution holds much more explanatory power than the standard normal for other types of data as well. Flexible parametric distributions are also shown to be important in McDonald, Hansen, and Theodossiou (2007).

Five Alternate Distributions

I will examine the probability density functions (from which it is easy to construct the likelihood function as the product of the densities) associated with each of the aforementioned five error distributions, followed by the results from each QMLE exercise in comparison to the standard OLS.

Normal (OLS) $$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$	Student's T (t) $$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$
Generalized Error Distribution (GED) $$\frac{pe^{-(\lvert u\rvert/\sigma)^p}}{2\sigma\Gamma(1/p)}$$	Inverse Hyperbolic Sine (IHS) $$\frac{ke^{-\left(\frac{k^2}{2}\left(\ln\left(u/\sigma+\sqrt{\theta^2+u^2/\sigma^2}\right)-(\lambda+\ln\theta)\right)^2\right)}}{\sqrt{2\pi\left(\theta^2+u^2/\sigma^2\right)\sigma^2}}$$
Generalized T (GT) $$\frac{p}{2\sigma q^{1/p}B(1/p,q)\left(1+\frac{\lvert\mu\rvert^p}{q\sigma^p}\right)^{q+1/p}}$$	
Skewed Generalized T (SGT) $$\frac{p}{\left[2\phi q^{1/p}B(1/p,q)\left(1+\frac{\lvert y-m\rvert^p}{q\left(1+sign(y-m)\lambda\right)^p\phi^p}\right)^{q+1/p}\right]}$$	

Parameters are defined for the t distribution as follows: υ is the degrees of freedom, and Γ() is the gamma function. For the GED, the lower the value on parameter p, the sharper the peaks of the distribution at the mean μ. The parameter θ is then a measure of scale. In the GT and SGT, k and λ ($-1<\lambda<1$) are kurtosis and skewness parameters respectively, with q as a measure of tail thickness. In SGT, the parameter *Φ* is a scalar, with B as the beta function, and m is the mode of the random variable y with other parameters defined as above (Theodossiou, 1998). Because of the high number of parameters, IHS, GT, and SGT are the most flexible, with SGT at the top of that list. Multiple parameters are especially important for allowing for differences in skewness and kurtosis. Skewness is accounted for with the IHS (not nested in SGT) and SGT functions, while the t, GED, and GT are symmetric.

Using these various density functions for the errors of various linear models, the objective function then becomes:

$$max_{(\beta,\xi)} \Sigma_{(t=1)}^{n} \ln f\left(Y_t - X_t \beta ; \xi\right),$$

where ξ represents the set of parameters of the distribution of choice (SGT, GED, etc) and β here and in the equation represents regression coefficients.

Now that the objective function and density functions have been introduced, I move on to applications of each of these types of partially adaptive estimators using maximization algorithms common to econometrics (simplex and gradient methods) and compare them to standard OLS estimates for each of my data sets. I also present likelihood ratio tests for each to gauge the statistical significance of the difference between log-likelihood values for nested models. I then discuss the results and present conclusions.

	OLS		t		GED		IHS		GT		SGT	
Log Wages	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Below Average Looks	**-.1488**	.0418	**-.1525**	.0391	**-.1488**	.0408	**-.1526**	.0391	**-.1575**	.0380	**-.1584**	.0374
Experience	**.0134**	.0012	**.0137**	.0012	**.0138**	.0012	**.0137**	.0012	**.0135**	.0012	**.0137**	.0012
Black	.0646	.0534	-.0960	.0544	-.0916	.0583	-.0947	.0550	-.0867	.0538	-.0835	.0559
Female	**-.4438**	.0309	**-.4344**	.0299	**-.4246**	.0290	**-.4331**	.0297	**-.4432**	.0296	**-.4413**	.0293
Married	.0522	.0319	.0343	.0308	.0369	.0304	.0344	.0306	.0358	.0311	.0348	.0307
Education (Years)	**.0703**	.0054	**.0706**	.0056	**.0716**	.0064	**.0708**	.0057	**.0698**	.0053	**.0701**	.0052
Constant	**.6724**	.0816	**.6757**	.0840	**.6595**	.0921	**.6704**	.0841	**.6870**	.0802	**.6776**	.0794
Log Likelihood	**-869.1485**		**-851.291**		**-857.094**		**-851.376**		**-849.977**		**-849.101**	

	OLS		t		GED		IHS		GT		SGT	
% Stud. Pass Math	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Avg Teacher Sal.	0.0001	0.0001	0.0001	0.0002	0.0001	0.0002	0.00006	0.0001	0.0001	0.0002	0.0000	0.0002
% Free Lunch Elig.	**-0.4025**	0.0385	**-0.4295**	0.0455	**-0.4300**	0.0457	**-0.4254**	0.0453	**-0.4307**	0.0454	**-0.4368**	0.0469
Staff//1000 Students	0.0081	0.0300	0.0037	0.0451	0.0078	0.0480	-0.0361	0.0357	0.0062	0.0476	0.0170	0.0511
Avg Teacher Ben.	**0.0009**	0.0003	**0.0010**	0.0004	**0.0010**	0.0004	0.0007	0.0003	**0.0010**	0.0004	**0.0012**	0.0004
Log Exp. Per Pupil	**9.0596**	3.9818	8.7791	5.9780	8.5237	6.3036	**18.0077**	4.90407	8.5996	6.2469	5.7927	6.7490
Log Student Enroll.	-0.2129	0.5593	-0.2903	0.6421	-0.2912	0.6712	-0.4616	0.6395	-0.2938	0.6599	-0.0164	0.6600
Constant	-35.9722	28.8024	-31.7380	42.9588	-30.4078	45.0611	**-100.1650**	37.344	-30.6931	44.6506	-10.0666	48.0121
Log Likelihood	**-2377.4**		**-2376.24**		**-2376.18**		**-2371.17**		**-2376.16**		**-2375.07**	

	OLS		t		GED		IHS		GT		SGT	
CEO Log Salary	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Log Sales	**.28065**	.03454	**.29254**	.02377	**.29966**	.01150	**.29031**	.02322	**.29230**	.02423	**.29072**	.02383
Return on Equity	**.01887**	.00405	**.01918**	.00272	**.01639**	.00068	**.01964**	.00264	**.01924**	.00266	**.01970**	.00259
Return on Stock	.00033	.00055	.00110	.00068	.00030	.00049	.00103	.00066	.00110	.00067	.00106	.00066
Industry	.01109	.07812	**.13642**	.05661	.01453	.01465	**.14950**	.05837	**.13736**	.05565	**.14915**	.05746
Finance	**.19038**	.08736	**.27771**	.07291	**.23158**	.02160	**.28479**	.07332	**.27829**	.07357	**.28476**	.07424
Constant	**4.23283**	.31324	**3.98982**	.21967	**4.07648**	.05241	**4.02336**	.21631	**3.99116**	.22206	**4.01182**	.21888
Log Likelihood	**-139.8787**		**-108.507**		**-113.076**		**-107.976**		**-108.489**		**-108.052**	

	OLS		t		GED		IHS		GT		SGT	
Cumulative GPA	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Lived On Campus	-0.1010	0.0771	-0.1010	0.0726	**-0.1232**	0.0503	-0.1161	0.0748	**-0.1073**	0.0482	-0.1124	0.6686
Age	0.0141	0.0238	0.0141	0.0255	0.0248	0.0134	0.003796	0.0296	0.01787	0.0117	0.014303	0.0932
High School GPA	**0.4414**	0.0972	**0.4410**	0.0936	**0.4819**	0.0620	**0.409425**	0.0898	**0.475287**	0.0515	0.521416	0.5873
ACT Score	0.0117	0.0108	0.0117	0.0105	0.0062	0.0068	0.010973	0.0096	0.004508	0.0059	-0.0028	0.0603
Worked 20+	-0.0813	0.0758	-0.0814	0.0766	**-0.1722**	0.0548	-0.0964	0.0768	**-0.1949**	0.0502	-0.1018	0.3985
BF/GF	0.0871	0.0578	0.0872	0.0532	**0.0852**	0.0414	**0.105396**	0.0511	**0.112447**	0.0407	0.087284	0.2105
Hours Cut/Week	**-0.0981**	0.0285	**-0.0981**	0.0263	**-0.0877**	0.0164	**-0.0928**	0.0239	**-0.0910**	0.0153	-0.0831	0.1589
Days/Wk Alcohol	0.0238	0.0232	0.0238	0.0236	**0.0424**	0.0131	0.018854	0.0229	**0.053301**	0.0114	0.02942	0.1741
Dad College Grad	0.0350	0.0623	0.0349	0.0599	0.0556	0.0400	0.051506	0.0564	0.028965	0.0360	0.010009	0.2235
Mom College Grad	-0.0201	0.0604	-0.0201	0.0569	-0.0348	0.0400	-0.026908	0.05405	-0.025218	0.03386	-0.034629	0.35779
Constant	**1.0188**	0.6679	1.0194	0.6982	0.7802	0.4102	1.353788	0.7755	**0.9747**	0.3542	1.11896	4.1278
Log Likelihood	**-37.3805**		**-37.3805**		**-31.9876**		**-36.4815**		**-31.5751**		**-27.8587**	

	OLS		t		GED		IHS		GT		SGT	
Crime Rate	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Log Police/1000	**0.4710**	0.1130	**0.4710**	0.0980	**0.5151**	0.0894	**0.4710**	0.0980	**0.5152**	0.0894	**0.5200**	0.0893
Log Enforc. Exp	0.1872	0.0987	**0.1871**	0.0936	**0.2184**	0.0811	**0.1871**	0.0936	**0.2184**	0.0811	**0.2176**	0.0814
Log Area	-0.0609	0.0392	-0.0609	0.0357	-0.0574	0.0332	-0.0609	0.0357	-0.0574	0.0332	-0.0555	0.0332
Log Pop. Density	**-0.1797**	0.0646	**-0.1797**	0.0572	-0.1831	0.0536	-0.1797	0.0572	-0.1831	0.0536	-0.1813	0.0535
Constant	**4.6915**	0.7739	**4.6915**	0.7582	**4.4658**	0.6652	**4.6915**	0.7582	**4.4663**	0.6652	**4.4441**	0.6676
Log Likelihood	**-1.5350**		**-1.5350**		**-0.5365**		**-1.5350**		**-0.5365**		**5.5278**	

	OLS		t		GED		IHS		GT		SGT	
Total Accidents	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Unemployment	**-1274.67**	191.39	**-1274.53**	203.76	**-1281.11**	183.10	**-1137.45**	94.67	**-1281.15**	183.10	**-1282.25**	189.84
Speed Law	-1193.36	720.20	-1192.85	691.16	-1196.61	671.66	379.36	380.34	-1196.43	671.66	-1208.59	677.12
Belt Law	**4917.24**	706.00	**4917.36**	700.42	**4969.68**	584.25	**4762.47**	399.07	**4969.56**	584.27	**4939.07**	596.52
Weekends	408.74	210.12	408.99	221.06	406.15	208.78	**346.90**	149.15	406.07	208.78	**441.82**	210.93
Constant	**44834.35**	3242.32	**44829.72**	3022.21	**44937.35**	2916.21	**43126.25**	2482.00	**44938.65**	2916.19	**44495.18**	2867.04
Log Likelihood	**-984.0789**		**-984.079**		**-983.689**		**-1016.45**		**-983.689**		**-983.178**	

Results

For all but two of the data sets, we see increasing improvement in likelihood across distribution choices as complexity and flexibility expand in nested models. However, for the remaining two, we see no improvements.

A summary of the likelihood ratio tests between OLS estimates and SGT estimates is included below. There seems to be a consistent improvement from OLS for almost all estimators, and certainly for those with at least two extra parameters. In addition, SGT appears to lead to a consistent improvement even among alternate specifications in most cases. I also note that gains in accuracy depend on the specific characteristics of the data being estimated. Figures of the residuals for some of the alternate specifications are provided in Appendix C.

Likelihood Ratio Tests—OLS

LR Test	Beauty	Math Score	CEO Salary	GPA	Crime	Traffic
T	35.715 ***	2.32	62.743***	0	.0000	-0.00026
GED	24.109 ***	2.44	53.606***	10.786 ***	1.9969	0.77974
IHS	35.545 ***	12.46	63.806***	1.798	.0000	-64.742***
GT	38.344 ***	2.48	62.780***	11.611 ***	1.9969	0.77974
SGT	40.095 ***	4.66	63.653***	19.044 ***	14.1256 ***	1.80174
Statistical significance given by: ***p<.01, **p<.05, *p<.1						

Likelihood Ratio Tests—SGT

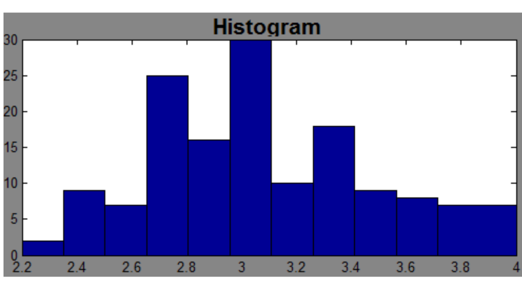
LR Test	Beauty	Math Score	CEO Salary	GPA	Crime	Traffic
T	4.38 **	2.34	0.91	19.044 ***	14.126 ***	1.802
GED	15.986 ***	2.22	10.048***	8.258 **	12.129 ***	1.022
IHS	4.55 **	-7.8	-0.152	17.246 ***	14.126 ***	-66.544***
GT	1.752	2.18	0.874	7.433 **	12.129***	1.022
Statistical significance given by: ***p<.01, **p<.05, *p<.1						

Discussion

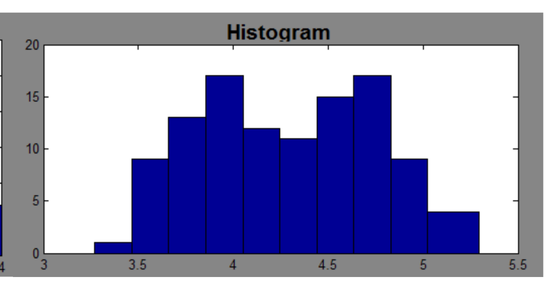
Notice from our likelihood ratio tests that there is a significant lack of improvement in likelihood in some regressions, the reasons for which vary.

GPA and Traffic

The models that have the most continuous, normally distributed independent variables see strong improvements in likelihood (Beauty, Crime, CEO Salary). The GPA regression saw no statistical increase in likelihood for any lower order specifications. The reason for this may be the small spread of the dependent variable (GPA) as it only covers between 2.2 and 4.0. This combined with the relatively low number of observations (141) in the presence of a high number of binary or ordered variables (7) make statistical inference difficult under any circumstance. We see this in the Traffic data set as well as all the independent variables in the regression are either binary or ordered. This undermines the ability of least squares to reliably predict the relationship between each independent variable and the dependent variable. The initial power of an ordinary least squares model acts as a starting point for more computationally complicated estimators. In other words, a poorly defined linear model will yield little valuable information or improvements in estimation *regardless* of the choice of approximating error distribution. Continuity and sufficient variation of the X's are as essential for choosing parameter values in maximum likelihood as they are in an OLS framework. As is noticeable from the histogram of the traffic dependent variable, the binary nature of the independent variables (as we see in these histograms) appears to have a bifurcating effect which divides out the “peak” of our distribution and changes the kurtosis of the residual probability density function away from its typical value of 3 common to the normal distribution. A similar pattern emerges in GPA, although the bifurcation of the dependent variable is not as pronounced.



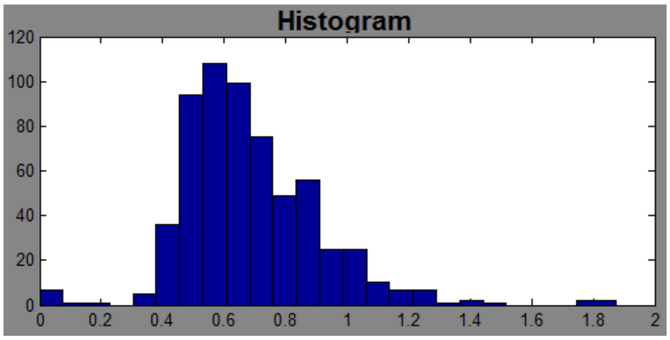
GPA: College GPA



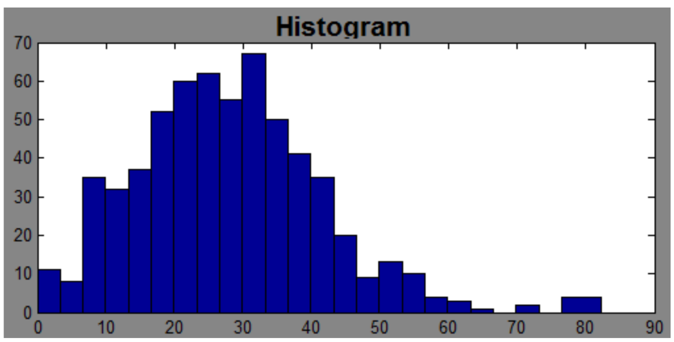
Traffic: Total Accidents

Math Proficiency

Unlike the continuity problem of the previous two data sets, the math proficiency data set has some interesting distribution characteristics to examine.



Math: Average Teacher Benefits



Math: Students Qualifying for Free Lunch

The large skewness across the math proficiency independent variables may add to the difficulty of calculating distribution parameters that can adequately fit the data. Skewed X's may place the residuals in an undefined skewness-kurtosis space, making flexibility of little use if the parameters cannot reach more efficient values. Flexible parameters seem to take into account kurtosis, but skewness becomes a problem for many of these distributions. A combination of kurtosis and skewness in the true residuals make it increasingly difficult to satisfy the needs of both with a parametric residual. A mixture residual model or kernel may perform better in these circumstances. In addition, this is a panel data set in which the multiple observations are constituted by several different yearly observations of the same school. The effect this has on the outcome is unclear.

Crime Rate

Our crime rate regression experiences neither of the problems of the above data sets. In fact, the reason there is no improvement in estimation for intermediate flexible distributions is that the logarithmic transformation of *all* variables in the regression leads to stronger OLS performance and mitigates outlier groups that skew distributions. In regressions where every variable has been transformed or where the variables are normally distributed, differences between OLS and MLE with flexible distributions are minimal. In effect, OLS parameters have neared an efficiency maximum with this data set. Even with these log transformations, there is still improvement in likelihood at the SGT level, making it meaningful to explore.

Comparison with SGT Results

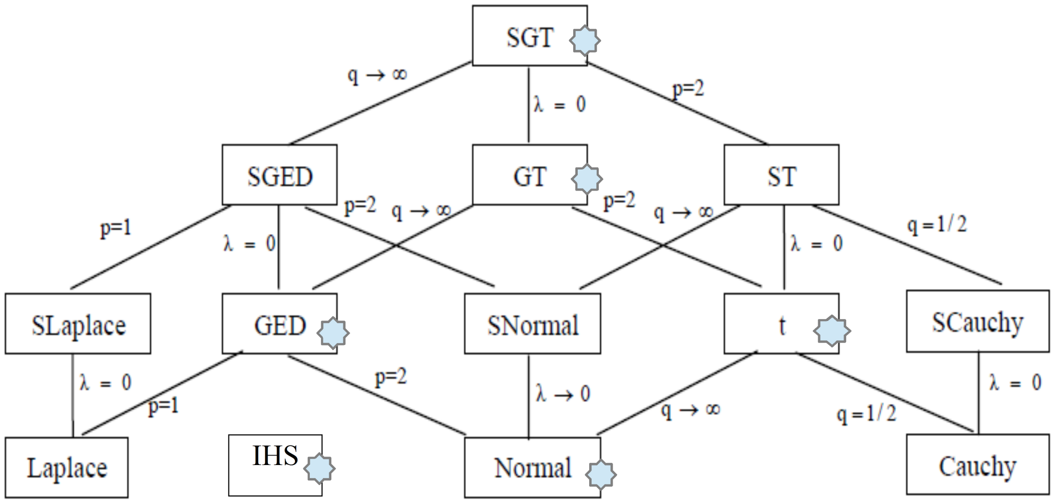
The purpose of the likelihood ratios tests of SGT results to others is an examination of parsimony. In the Beauty regression, SGT was significantly different from GED, but not from the GT, which is simpler than SGT and yields estimates that are just as strong. The CEO Salary regression elicits consistent improvements in likelihood, but there is no significant difference between SGT and any lower order distribution other than GED. Due to the complexity of using SGT compared to a Student's t, one would benefit from choosing the simpler t rather than the unnecessarily complex SGT or GT. This is an important consideration when choosing the distribution that will allow for the greatest gains in accuracy without sacrificing computational simplicity. In the Crime Rate regression, SGT was the only specification that yielded better likelihood results, in which case a lower level specification would not yield similar outcomes. From this analysis, we see that it is important to test for likelihood across flexible distributions as likelihood improvements are contingent upon specific data characteristics and vary widely.

Conclusion

It is clear from this analysis that for models with specific types of data or model specification, improvements from increased residual flexibility may be minimal, especially when the the error distribution is already approximately normal. However, for the majority of moderately well-defined linear models, there appears to be a marked improvements in log likelihood, at least from our examples. The choice of model is really left to the discretion of the researcher as to where their particular interests or priorities lie and the characteristics of their data. The SGT distribution appears to lead to improvements across data types, but this depends on further research to demonstrate more assuredly. However, the principle of parsimony applies well here: choice of a model should be limited to the simplest version required to achieve efficiency and accuracy. In some

instances, the GT or IHS (which is not nested in SGT, but performs in a similar fashion to SGT) model is perfectly suitable, while other times, a t distribution or GED performs just as well. To the extent that the results of a particular estimation method influence public policy and hold large consequences, it would be wise for researchers to examine several alternate specifications and their relative strengths to ascertain the true effects of the variables being scrutinized. Improvements in computational power allow more rigorous analysis—including alternate error distribution specifications—to be more accessible to policy researchers, academics, and interested parties. This type of analysis should be complementary to typical empirical papers, with extra emphasis placed upon these specifications. Such a necessity illustrates possible gains to be made in the field of econometrics in general.

Appendix A
Skewed Generalized t Distribution Family



Note: From McDonald, Hansen, and Theodossiou (2007). The distributions examined is this paper are highlighted by .

Appendix B

Summary Statistics

BEAUTY (from beauty.dta)

Variable	Obs	Mean	Std. Dev.	Min	Max
Log Wage	1260	1.6588	0.5945075	0.0198026	4.353113
Education	1260	12.56349	2.624489	5	17
Experience	1260	18.20635	11.96349	0	48
Black	1260	0.0738095	0.2615645	0	1
Female	1260	0.3460317	0.4758923	0	1
Married	1260	0.6912698	0.462153	0	1
Below Average Looks	1260	0.1230159	0.3285858	0	1

MATH PROFICIENCY (from mathpnl.dta)

Variable	Obs	Mean	Std. Dev.	Min	Max
Percent of Students Passing Math	1631	38.07903	14.00702	0	84.1
Average Salary of Teachers	1631	31395.39	6346.631	0	76412
Percent Free Lunch Eligible	1631	25.96658	14.65548	0	83.3
Staff per 1000 Students	1631	102.1671	22.67785	0	384.1
Average Teacher Benefits	1631	7151.632	2312.054	0	18754
Log Expenditure per Pupil	1631	8.378605	0.2063228	7.044905	9.255409
Log Student Enrollment	1631	7.471395	1.015308	3.912023	12.11807

CEO SALARY (from ceaosal1.dta)

Variable	Obs	Mean	Std. Dev.	Min	Max
Log Salary	209	6.950386	0.5663741	5.407172	9.603868
Log Sales	209	8.292265	1.013161	5.165928	11.48914
Return on Equity	209	17.18421	8.518509	0.5	56.3
Return on Stock	209	61.80383	68.17705	-58	418
Heavy Industry	209	0.3205742	0.4678178	0	1
Finance	209	0.2200957	0.4153057	0	1

GPA (from gpa2.dta)

Variable	Obs	Mean	Std. Dev.	Min	Max
College GPA	141	3.05674	0.37231	2.2	4
Lived on Campus	141	0.170213	0.377159	0	1
Age	141	20.8865	1.27106	1.9	3
High School GPA	141	3.40213	0.319926	2.4	4
ACT Score	141	2.4156	0.284425	1.6	3.3
Worked 20 Hours per Week	141	0.170213	0.377159	0	1
Boyfriend/Girlfriend	141	0.475177	0.501164	0	1
Classed Skipped	141	1.07624	1.08888	0	5
Days/Week Consumed Alcohol	141	1.90106	1.3747	0	7
Father College Graduate	141	0.588652	4.93832	0	1
Mother College Graduate	141	0.539007	5.00253	0	1

CRIME RATES (from crime2.dta)

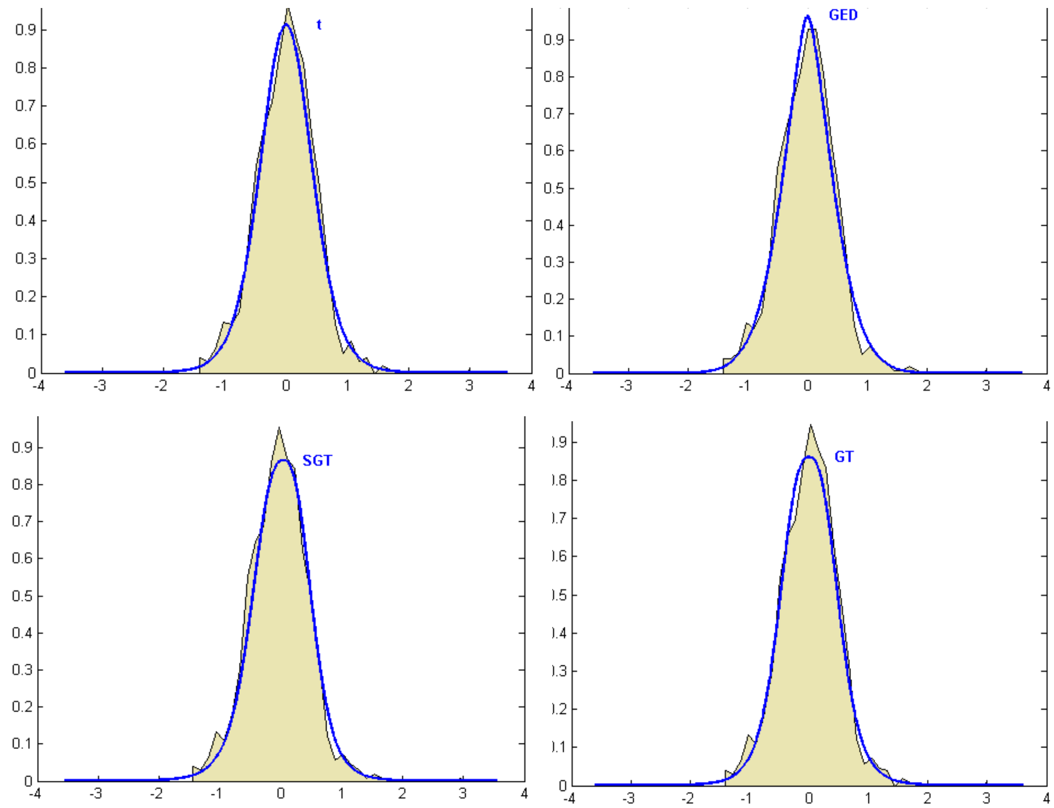
Variable	Obs	Mean	Std. Dev.	Min	Max
Log Crime Rate (per 1000)	92	4.57239	0.283785	3.91241	5.18971
Log Population Density	92	8.31728	0.65474	6.55649	9.71416
Log Area	92	4.35583	0.961325	2.56495	6.40357
Log Law Enforcement Exp	92	6.81724	0.308913	5.93368	7.7242
Log Police per Thousand	92	0.774195	0.274798	0.249933	1.53023

TRAFFIC ACCIDENTS (from traffic2.dta)

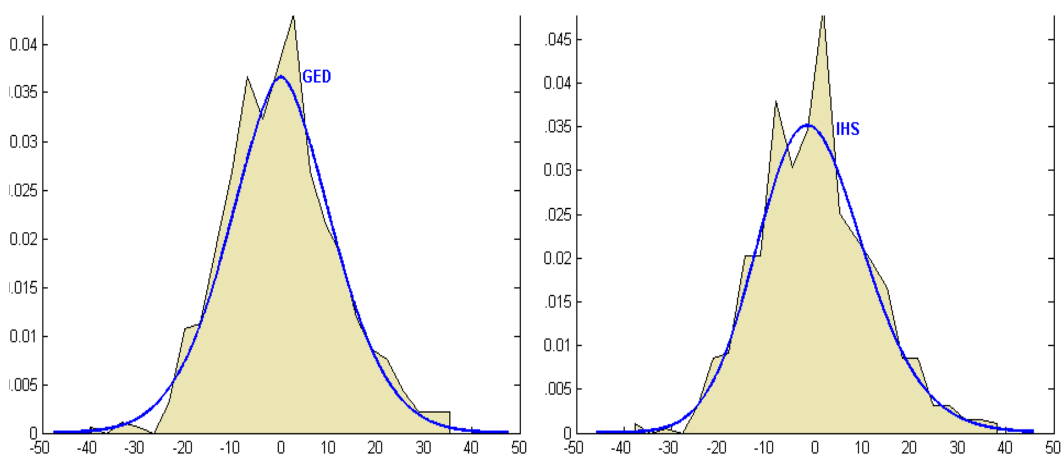
Variable	Obs	Mean	Std. Dev.	Min	Max
Total Accidents	108	42831.26	4608.328	32699	52971
Unemployment Rate	108	7.200926	1.790134	4.3	11.9
Speed Law Enforced	108	0.2962963	0.4587521	0	1
Seatbelt Law Enforced	108	0.4444444	0.4992206	0	1
Weekend Days per Month	108	13.07407	1.011187	12	15

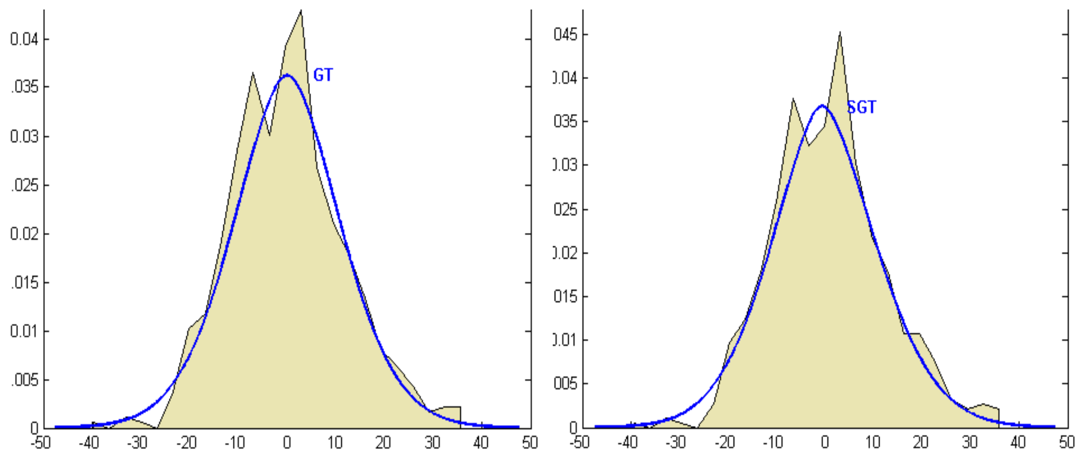
Appendix C—Residual Graphs

BEAUTY

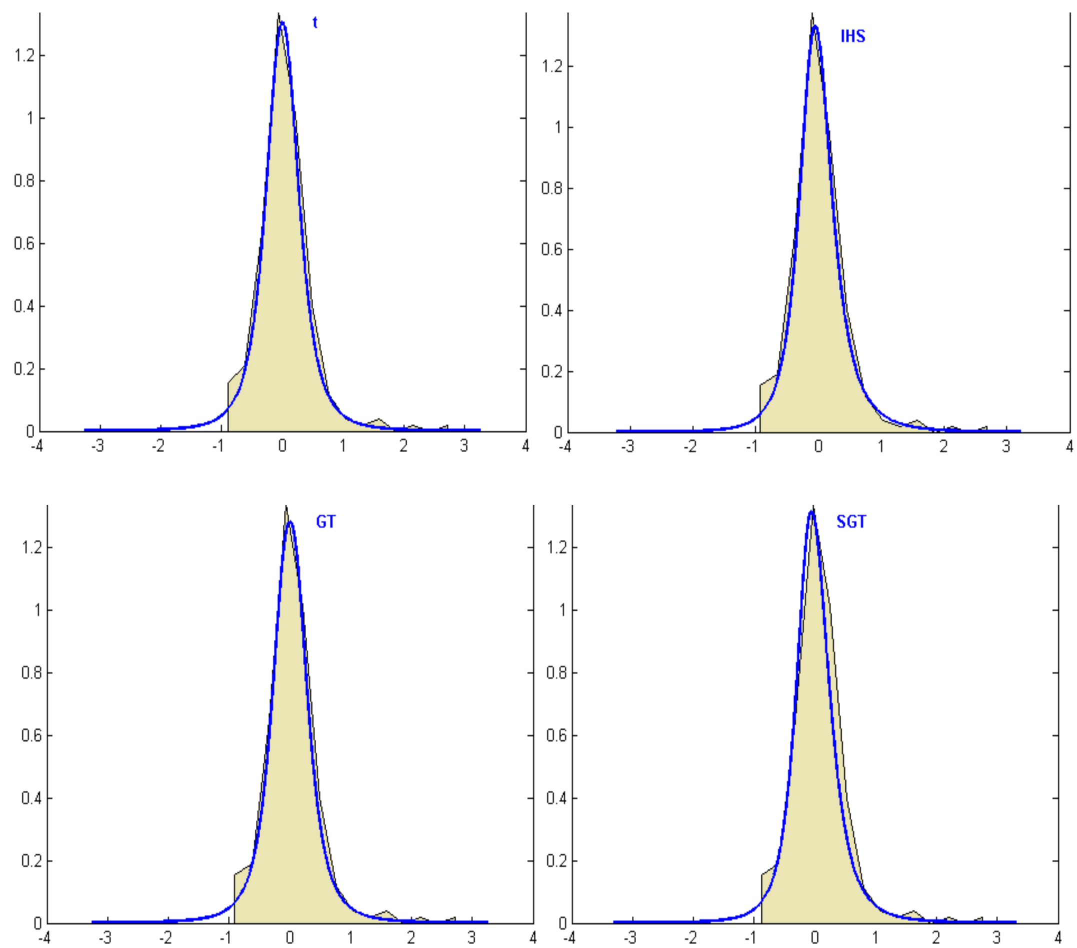


MATH PROFICIENCY

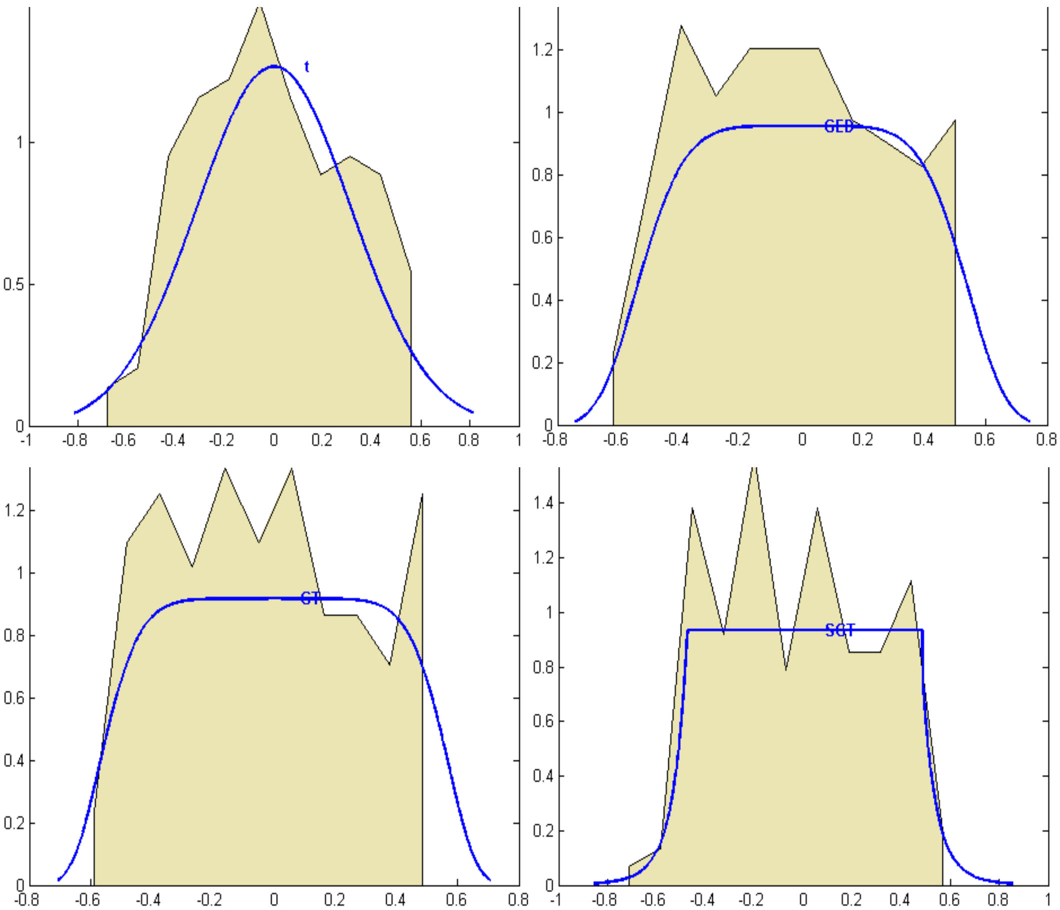




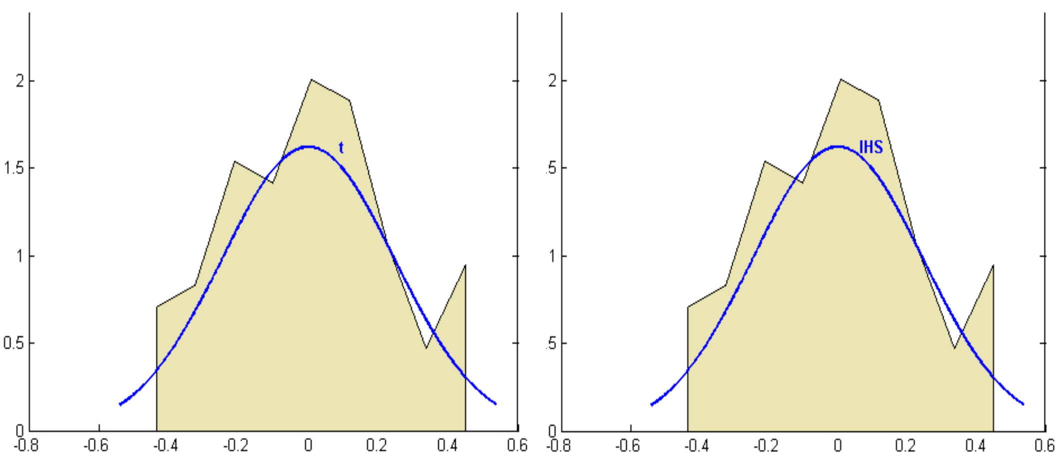
CEO SALARY

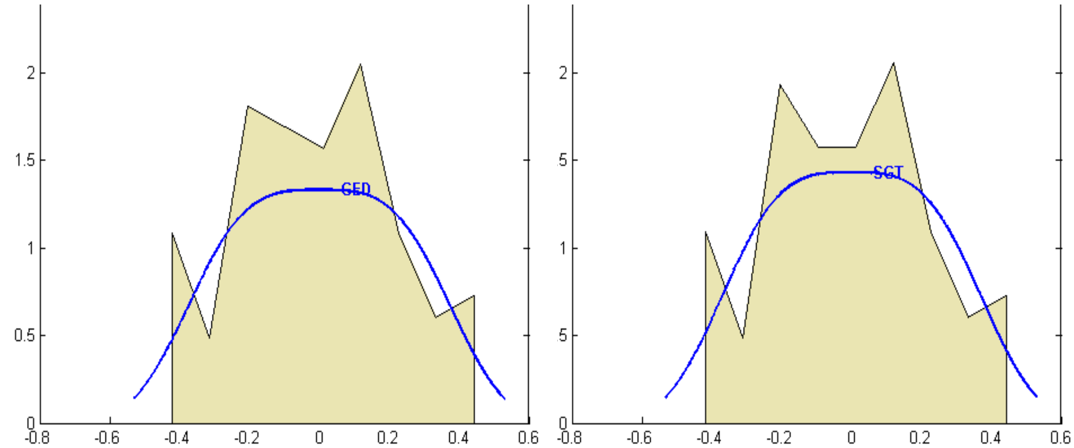


GPA

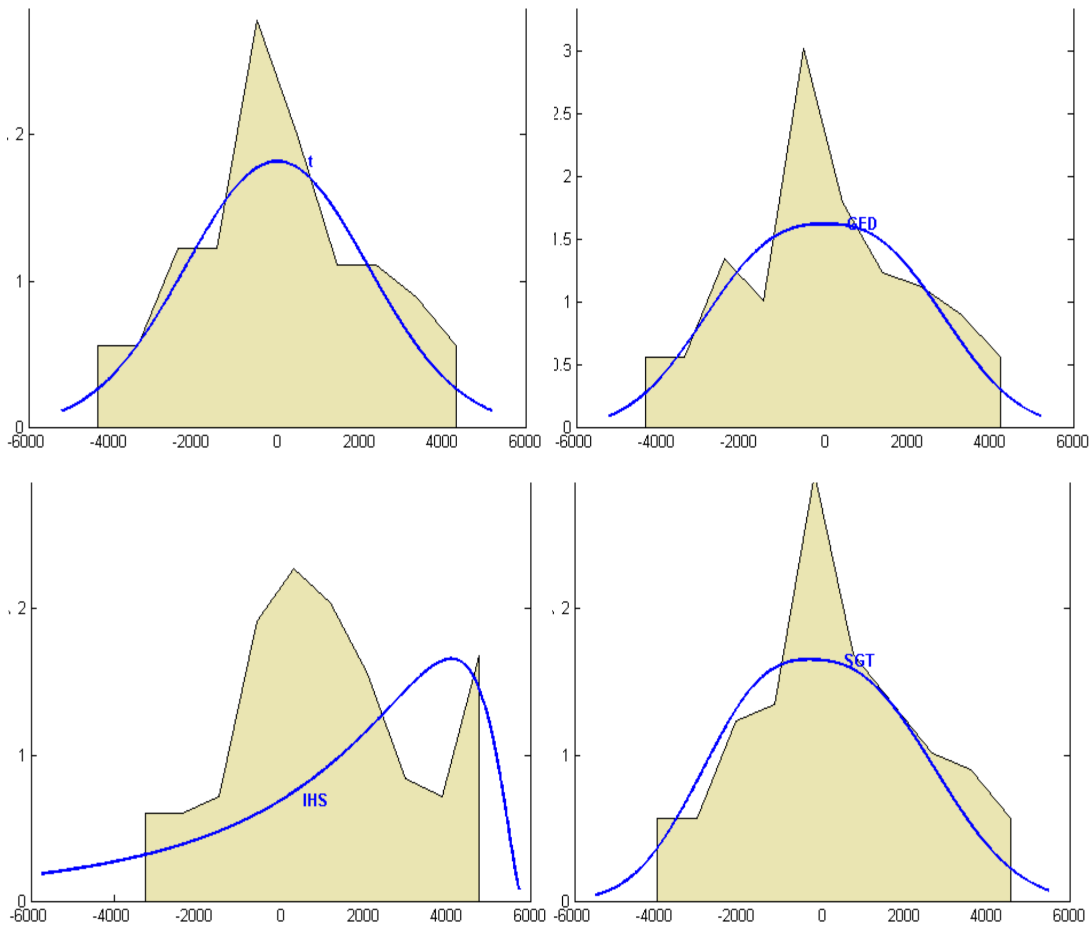


CRIME RATES




GED
SGT
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6

TRAFFIC


t
GED
IHS
SGT
-6000
-4000
-2000
0
2000
4000
6000

References

Box, G.E.P., and Tiao, George C., "A Bayesian Approach to the Importance of Assumptions Applied to the Comparison of Variances," *Biometrika*, 51 (1964), pp.153-167.

Efromovich, Sam, "Estimation of the Density of Regression Errors," *The Annals of Statistics,* Vol. 33, No 5 (Oct, 2005), pp 2194-2227.

Hansen, Bruce E., "Autoregressive Conditional Density Estimation," *International Economic Review* Vol 35, No 3, (Aug. 1994), pp 705-730.

McDonald, James B., "Some Generalized Functions for the Size Distribution of Income." *Econometrica* Vol 52, No 3 (1984) pp 647-665.

McDonald, James B., Hansen, Christian B., and Theodossiou, Panayiotis, "Some Flexible Parametric Models for Partially Adaptive Estimators of Econometric Models," *Economics.* (2007).

McDonald, James B., and Newey, Whitney. "Partially Adaptive Estimation of Regression Models via the Generalized T Distribution." *Econometric Theory*, Vol 4 (1988), pp 428-457.

McDonald, James B., and White, S. B., A Comparison of Some Robust, Adaptive, and Partially Adaptive Estimators of Regression Models. *Econometric Reviews* Vol 12 No 1 (1993), 103-124.

Theodossiou, Panayiotis, "Financial data and the Skewed Generalized T Distribution," *Management Science,* Vol 44. No 12 (Dec. 1998), pp. 1650-1661.

Theodossiou, Panayiotis, "Skewed Generalized Error Distribution of Financial Assets and Option Pricing," Rutgers University School of Business Working Paper Series, (2000).

Zeckhauser, Richard, and Thompson, Mark, "Linear Regression with Non-Normal Error Terms," *The Review of Economics and Statistics*, Vol 52. No 3 (Aug., 1970), pp. 280-286.