How Non-Normal is US Output?

Reiner Franke [a,*]

First Draft, January 2013

[a]University of Kiel, Germany

Abstract

Several studies have recently rejected the common hypothesis that aggregate output is normally distributed, a finding that should also have a bearing on DSGE modelling. The present paper reconsiders this issue for quarterly US output data. To this end five test statistics are adopted, among them the shape parameter of the exponential power distribution (EPD), the two polar values of which constitute the normal distribution and the Laplace distribution with its fatter tails, respectively. The main results are: (1) Evidence of non-normality of the output gap disappears once it is controlled for the high serial correlation. (2) The non-normality results in the literature concerning the growth rates can be explained by normality in two subsamples once a structural break is taken into account. (3) The only way to detect non-normality in the subsamples is the estimation of the shape of EPD. (4) Normality cannot be rejected in the Great Inflation period and the Laplacian cannot be rejected for the period of the Great Moderation.

JEL classification: C 22, E 32.

Keywords: Spurious non-normality, exponential power distribution, generalizations of Jarque-Bera, Anderson-Darling test, Monte Carlo experiments.

1. Introduction

Macroeconomic variables or the shock processes that drive them are commonly considered to be largely compatible with normal distributions. This supposition is not just a semantic issue, for example in assessing whether or not our economic system, occasional crises notwithstanding, is mainly well-behaved and the risk from the random perturbations in it can be satisfactorily controlled. It is also an important question in the academic field, where the flourishing business of the estimation of DSGE models with

* *Email address:* `franke@uni-bremen.de`.

its present predominance of likelihood techniques heavily rests on the assumption of normally distributed innovations.

There have nevertheless been always some doubts about the prevalent view that takes normality for granted, in the first instance with respect to output and employment variables (as far as the empirical side is concerned, the following will refer to US data). Independently of discussions among heterodox theorists (like Blatt, 1983), concerns about asymmetry were already raised some thirty years ago. The possibly best-known contribution at that time was DeLong and Summers (1985) who, however, did not get much evidence for this. More recently, the tests by Bai and Ng (2005) failed to establish significant support of skewness or excess kurtosis. In contrast to these negative results, Christiano (2007) obtained significant excess kurtosis in the residuals of an unconstrained estimated vector autoregression. Interestingly, on the other hand, this did not prove sufficient to distort the Bayesian analyses that use the normal likelihood.

Other work finds indications of non-normality, too, and tries to take them into account in their theoretical models. De Grauwe (2012) concentrates on the kurtosis of output and emphasizes that, despite the normal shock processes, the nonlinear mechanisms in his model are able to reproduce this feature with great success.[1] While De Grauwe's discussion remains more informal, Ruge-Murcia (2012) rigorously estimates the third-order approximation of a DSGE model by the simulated method of moments, an approach that need not rely on normality assumptions. Within this framework he can reject the null that productivity innovations are normally distributed in favour of the alternative that they follow an asymmetric Skew normal distribution.[2]

Besides studying skewness and kurtosis, there is another approach that assesses deviations of the empirical distribution in its entirety from normality. To this end, the analysis refers to the class of the exponential power (or Subbotin) distributions. Their general shape is governed by a parameter b, where $b=2$ yields the normal distribution and lower values give rise to progressively fatter tails. An alternative benchmark is then $b=1$, at which the Laplace distribution prevails.[3] Several papers by Fagiolo et al. (2007, 2008, 2009) claim it a universal phenomenon for output growth rates that estimates of b are so low that normality has to be rejected. Ascari et al. (2012) even raise this into the category of a stylized fact, i.e., a standard that macroeconomic models should seek to meet. Checking this with calibrated versions of a Real Business Cycle and a New-Keynesian model, they find out that the former can replicate this type of fat tails exogenously but

[1] The strong nonlinearities originate with the agents' switching between different types of boundedly rational expectations, which he puts forward as a pronounced alternative to DSGE modelling.
[2] The moments themselves on which the model is estimated include the third-order moments of hours worked and of the consumption and investment growth rates; see Ruge-Murcia (2012, p. 931, Table 10).
[3] It plays a prominent role in cross-sectional distributions of firm characteristics such as growth rates and profit rates; see, e.g., Alfarano et al. (2012).

not endogenously, and the latter neither endogenously nor exogenously. Thus, this work raises a serious criticism against the current practice of DSGE modelling.

The present paper sets in with the observation for US quarterly output data that all of the studies revealing non-normality are based on samples covering 40 years and more. Whereas such a sample size is certainly desirable from an econometric point of view, so many things have changed historically over this span of time that also the economy may be suspected of having shifted from one regime to another. Somewhat strangely, this possible problem is not touched upon in any of the studies. After all, the distinction between the two periods of the Great Inflation and the Great Moderation is a well-known topic in macroeconomics. In addition and more specifically, McConnel and Perez-Quiros (2000) provided firm statistical evidence of a structural break in the volatility of output growth around 1984.

An obvious question thus arising is whether the features of non-normality can also be detected in the two subsamples before and after the structural break, even though with weaker rejections of normality because of the smaller samples. Regarding the long sample period, one may furthermore ask if the non-normality, if it is found to prevail at all, is perhaps spurious in the sense that it could be alternatively explained by the pooling of two samples which are characterized by two normal distributions with—according to the structural break—different dispersion. Apart from this, it may be checked whether the different test statistics used in the aforementioned literature all come to similar conclusions. In short, it is time that the issue of non-normality, even if confined to the levels and growth rates of US output, be subjected to a careful reconsideration. This is the purpose of the present paper.

The investigation is organized as follows. Section 2 presents five test statistics to detect non-normality. On the one hand, these are the popular Jarque-Bera test and two generalizations that take account of the serial correlation in the data. The other two tests are concerned with the overall shape of the distributions, namely, the Anderson-Darling test and the shape parameter of the exponential power distribution, which has already been mentioned. The remaining sections deal with the quarterly data of US output, where we focus on the growth rates of GDP and the output of the firm sector after the CBO output gap will have been found to offer less prospects for non-normality. Three different samples are moreover considered: the full sample of 47 years from 1960 on, and its decomposition into two almost equally long subsamples.

Section 3 applies the five tests to the empirical data and to the residuals from suitable autoregressions. While the conclusions drawn here are based on asymptotic theory, Section 4 uses Monte Carlo experiments to learn more about the small sample properties. Section 5 takes up the idea from above and indeed finds out that the non-normality results in the full sample could be satisfactorily explained by two estimated AR(p) processes over the two subsamples, the two normally distributed innovations of which have distinctly different variances.

As we will learn that the only feature of possible non-normality in a subsample is a low estimate of the shape parameter of the exponential power distribution, Section 6 puts forward the alternative hypothesis that the growth rates follow a Laplace distribution. This gives rise to the most pronounced result of our work: the Laplace can be safely rejected in favour of normality in one subsample, and it cannot be rejected in the other.

Lastly, Section 7 is concerned with the precision with which the shape parameter can be estimated. This issue is of particular relevance for models that may have the ambition to reproduce this aspect of normality or non-normality. Section 8 concludes, and an appendix contains several technical details.

2. Test statistics to detect non-normality

There are various fields in applied macroeconomic research where it is of interest whether a given realization of a stochastic process could have been obtained from a normal distribution. The most common approach to checking the normality of the marginal distribution of the data are procedures that test whether the third and fourth moments coincide with those of the normal distribution.[4] Accordingly, let a stationary univariate time series $\{x_t\}_{t=1}^T$ of length T be given with mean $\bar{x}$ and estimated standard deviation $\hat{\sigma}$. Its skewness S and kurtosis K are estimated as

$$\begin{aligned} \widehat{S} &= \hat{\mu}_3 / \hat{\sigma}^3 \\ \widehat{K} &= \hat{\mu}_4 / \hat{\sigma}^4 \\ \hat{\mu}_k &= (1/T) \sum_{t=1}^{T} (x_t - \bar{x})^k , \quad k = 2,3,4 \qquad (\text{hence } \hat{\mu}_2 = \hat{\sigma}^2) \end{aligned} \tag{1}$$

The normal distribution has $S = 0$ and $K = 3$. When $\widehat{K}$ is 'sufficiently' larger than 3, the distribution of x_t is said to exhibit excess kurtosis, or to have fat tails. It is also well-known that reliable estimates of the kurtosis require a fairly large number of observations, larger than the typical sample size of macroeconomic quarterly data. For this reason and in order to limit the discussion, we focus on tests of the joint hypothesis $S = 0$ and $K = 3$. Because of its simplicity the probably most popular test statistic is that of Jarque and Bera (1980),

$$JB \;=\; T \left[\frac{\widehat{S}^2}{6} + \frac{(\widehat{K}-3)^2}{24} \right] \;=\; T \left[\frac{\hat{\mu}_3^2}{6\,\hat{\mu}_2^3} + \frac{(\hat{\mu}_4 - 3\,\hat{\mu}_2^2)^2}{24\,\hat{\mu}_2^4} \right] \tag{2}$$

(the second expression is added for a better comparison with the generalized statistic below). If the random variable x_t is iid and normally distributed, JB is χ^2-distributed with two degrees of freedom. Thus normality will be rejected at a 5% significance level if

[4] There are nevertheless other probability distributions with the same two moments; see distribution S4 in Bai and Ng (2005, p. 60), the skewness and kurtosis of which are reported in the first four tables of this paper.

JB exceeds 5.99. (Throughout the paper, "significance" statements will be based on this level.)

There are two problems with this straightforward rule. First, the small-sample properties of the test are different from the asymptotic result. Second and more seriously, most macroeconomic time series data violate the prerequisite of iid. In the presence of serial correlation, however, the true asymptotic variances of S and $(K-3)$ are no longer consistently estimated by the denominators of *JB*, which implies that even asymptotically the rejection probabilities deviate from the desired nominal levels.

We consider two approaches to remedy these distortions, both of which require no deep assumptions on the true data generation process. [5] The first approach, which is borrowed from Lobato and Velasco (2004), modifies the variances 6 and 24 in (2) directly by taking the autocovariances of the series into account. The main correction terms are here $F^{(3)}$ and $F^{(4)}$ for the skewness and kurtosis statistics, respectively, defined as $F^{(k)} = \sum_{j=-\infty}^{\infty} \gamma(j)^k$ with respect to the population autocovariances $\gamma(j)$ of order j and $k = 3, 4$. For finite samples these sums can be estimated as

$$\begin{aligned} \widehat{F}^{(k)} &= \sum_{j=1-T}^{T-1} \hat{\gamma}(j) \, [\, \hat{\gamma}(j) + \hat{\gamma}(T-|j|)\,]^{k-1} \\ \hat{\gamma}(j) &= (1/T) \sum_{t=1}^{T-|j|} (x_t - \bar{x})\,(x_{t+|j|} - \bar{x}) \end{aligned} \tag{3}$$

($\hat{\gamma}(T)$ is set equal to zero). Lobato and Velasco (2004, p. 676) establish that asymptotically, for weakly dependent processes and under the null hypothesis of normality, their generalized Jarque-Bera statistic (*GJB*) is again $\chi^2(2)$ distributed, [6]

$$GJB = T \left[\frac{\hat{\mu}_3^2}{6\,\widehat{F}^{(3)}} + \frac{(\hat{\mu}_4 - 3\,\hat{\mu}_2^2)^2}{24\,\widehat{F}^{(4)}} \right] \xrightarrow{d} \chi^2(2) \tag{4}$$

This specification is indeed meaningful since $\widehat{F}^{(3)}$ and $\widehat{F}^{(4)}$ are ensured to be positive (p. 678). From a comparison of (2) and (4) it is furthermore easily seen that asymptotically *GJB* reduces to *JB* if the stochastic process is iid, since in this case $\hat{\gamma}(j) \to 0$ for all $j \neq 0$ in (3) and $\hat{\gamma}(0) = \hat{\sigma}^2 = \hat{\mu}_2$. With positive serial correlation in the first few lags of a time series, however, the denominator in *GJB* will be larger than in *JB*, so that *GJB* will fall short of *JB* and the chances of rejecting normality would decrease.

Also a converse conclusion holds true: under certain regularity conditions (technically requiring finite moments up to the sixteenth order) *GJB* diverges to infinity if the null

[5] Vavra and Psaradakis (2001) provide a third and more ambitious generalization of *JB*, which is based on smoothed quantiles and also uses more robust measures of the skewness and kurtosis. There is nevertheless no simple rule of thumb for the optimal specification of these quantiles (p. 14), which is the reason why we better wait for additional experience with this method.

[6] The square for $\hat{\mu}_2$ in the kurtosis expression is missing in the definition of their statistics *SK* and *G* (Lobato and Velasco, 2004, pp. 674f), which is a typo.

is violated, i.e., if for the population moments $\mu_3 \neq 0$ or $\mu_4 \neq 3\,\mu_2^2$. Hence normality will be rejected with a probability tending to one as $T \to \infty$.

The second approach to correct the asymptotic variance of *JB* was proposed by Bai and Ng (2005). This generalization is less closely related to *JB* than *GJB*. Rather, the testing procedure is more similar to a GMM test of overidentifying restrictions, though some subtle differences still remain. In detail, define

$$z_t \;=\; \begin{bmatrix} x_t - \bar{x} \\ (x_t - \bar{x})^2 - \hat{\sigma}^2 \\ (x_t - \bar{x})^3 \\ (x_t - \bar{x})^4 - 3\hat{\sigma}^4 \end{bmatrix}, \qquad \bar{z} \;=\; (1/T)\sum_{t=1}^{T} z_t$$

$$y_T \;=\; \begin{bmatrix} (1/\sqrt{T})\;\sum_t (x_t - \bar{x})^3 \\ (1/\sqrt{T})\;\sum_t [\,(x_t - \bar{x})^4 - 3\hat{\sigma}^4\,] \end{bmatrix}$$

$$\widehat{\alpha} \;=\; \begin{bmatrix} -3\hat{\sigma}^2 & 0 & 1 & 0 \\ 0 & -6\hat{\sigma}^2 & 0 & 1 \end{bmatrix}$$

$$\widehat{\Phi} \;=\; \Gamma_0 \;+\; \sum_{j=1}^{p} (1 - \frac{j}{p+1})\,(\Gamma_j \;+\; \Gamma_j')\;, \qquad p \;=\; [T^{1/4}]$$

$$\Gamma_j \;=\; (1/T)\sum_{t=1}^{T-j} (z_t - \bar{z})(z_{t+j} - \bar{z})'\;, \qquad j = 0, 1, \ldots, p$$

The (4×4) matrix $\widehat{\Phi}$ is a Newey-West estimator of the long-run covariance matrix of z_t that uses the linearly declining weights of the Bartlett kernel, where the maximal lag length $[T^{1/4}]$ (the smallest integer greater than or equal to $T^{1/4}$) is determined by a common rule of thumb (see Greene, 2002, p. 267, fn 10).[7] On this basis, Bai and Ng (2005, p. 52) specify a statistic, which we denote *BN*, and (for any consistent estimator of the covariance matrix of z_t) demonstrate that its distribution converges to $\chi^2(2)$:

$$BN \;=\; y_T'\,(\widehat{\alpha}\,\widehat{\Phi}\,\widehat{\alpha}')^{-1}\,y_T \;\xrightarrow{d}\; \chi^2(2) \tag{5}$$

The tests so far were concerned with the values of the third and fourth moments as they would be implied by the normal distribution. We are now turning to two tests that seek to take account of the shape of the entire distribution. The first one is the Anderson-Darling test. It is based on an evaluation of the squared differences between the hypothesized and the empirical distribution, which however places more weight on the observations in the tails of the distribution. In this respect it follows a similar idea to the kurtosis.

[7] Lobato and Velasco (2004, p. 675) emphasize that their statistic (4) does not require the introduction of a kernel function or such a user-chosen number.

Generally, if $F = F(x)$ is the hypothesized distribution and $F_T = F_T(x)$ the empirical cumulative distribution function, their distance is measured as

$$T \int_{-\infty}^{\infty} \frac{[F_T(x) - F(x)]^2}{F(x)\,[1-F(x)]}\, dF(x)$$

To apply this concept to the normal distribution and finite samples, the data must first be transformed into the standardized values,

$$y_t \;=\; (x_t - \bar{x})\,/\,\hat{\sigma}\;, \qquad\qquad t = 1,\;\ldots\;,T \tag{6}$$

and arranged in ascending order, $y_1 \leq y_2 \leq \;\ldots\; \leq y_T$. These y_i enter the standard normal cumulative distribution function, $\Phi = \Phi(y)$, to set up the Anderson-Darling test statistic AD.[8] Together with a simple rule to reject normality at the 5% significance level, the prescription reads,

$$\begin{aligned} A^2 &= -T \;-\; \frac{1}{T} \sum_{i=1}^{T} \{\,(2i-1)\,\ln\Phi(y_i) \;+\; [2(T-i)+1]\,\ln[1-\Phi(y_i)]\,\} \\ AD &= A^2 \left[\,1 + \frac{4}{T} - \frac{25}{T^2}\,\right]\;; \qquad \text{normality rejected if }\; AD > 0.787 \end{aligned} \tag{7}$$

While the Anderson-Darling test statistic tells us when to reject normality, it does not indicate along which dimension the empirical distribution may differ most from normality. In this respect a parametric approach is more useful that includes the normal distribution as a benchmark case; in particular, when it also provides information about whether or to what extent the empirical distribution may exhibit fat-tail behaviour. A flexible statistical tool for this is the family of the exponential power (EP) densities, or the Subbotin density functions,

$$f(x;b,a,m) \;=\; \frac{1}{2a\,b^{1/2}\,\Gamma(1+1/b)} \;\; \exp\left\{\,\frac{1}{b}\left|\frac{x-m}{a}\right|^b\,\right\} \tag{8}$$

where $\Gamma(\cdot)$ is the Gamma function.[9] The three parameters identifying the EP distributions are the location parameter m, the scale parameter a, and the shape parameter b. The latter determines the fatness of the tails and is therefore of prime importance to us. The bell shape of the normal density arises with $b = 2$, while for $b \to \infty$ the distribution tends towards the uniform distribution with support $[-a, a]$. On the other hand, as b decreases from the Gaussian benchmark, the shoulders of the distribution become smaller and the tails become fatter. The benchmark of practical concern in this direction

[8] It is well-known that $\Phi(\cdot)$ has no closed-form analytical expression. The details of its numerical approximation are given in the appendix.

[9] More general versions of (8) can also account for asymmetries; see, e.g., Botazzi and Secchi (2008), or Zhu and Zinde-Walsh (2009). We neglect this extension since skewness will not be much of a problem for us, the convenient moment-based estimation approach of eq. (9) below would no longer be applicable to two shape parameters b_1 and b_2, and we feel that our samples are typically too small for reliable results.

is $b = 1$, which, in a semi-log diagram, yields the tent shape of the Laplacian density (we will encounter it in Section 6, Figure 1, further below).

Since b is a parameter characterizing the global shape of the distribution, it can be expected that it will also be a more robust measure of the fatness of tails than the kurtosis. Theoretically, the kurtosis implied by an EP distribution is given by $K = K(b) = \Gamma(1/b)\,\Gamma(5/b)\,/\,[\Gamma(3/b)]^2$ (Chiodi, 1995, Section 2). Even if one prefers to refer to the kurtosis as a more familiar measure of fatness, this relationship may be used as a check of the direct empirical calculation of K. To get a first impression of its order of magnitude, *vis-à-vis* $K = 3$ for the normal distribution, the kurtosis of the Laplace distribution rises to $K = K(1) = 6$.

There are several likelihood methods to estimate the parameters of an EP distribution.[10] More convenient for us is a moment matching procedure proposed by Mineo (1994, 2003). Based on a generalized index of kurtosis as it can be derived for EP distributions, it permits an isolated estimation of b. For sample sizes of 100 or 200 observations it also appears to give the most accurate results (Mineo, 2003, p. 118). The approach does not require the optimization of an objective function but only the solution of an implicit equation in b, which is indicated by the exclamation mark in the first equation:

$$\frac{\hat{\sigma}}{\hat{d}} \;\overset{!}{=}\; \frac{\sqrt{\Gamma(3/b)\;\;\Gamma(1/b)}}{\Gamma(2/b)}\,, \qquad \hat{d} \;=\; (1/T)\sum_{t=1}^{T} |x_t - \bar{x}|$$
$$\text{iid normality rejected at 5\% level if} \quad \hat{b} \;<\; \begin{cases} 1.578 & \text{for } T = 190 \\ 1.456 & \text{for } T = \;\;95 \end{cases} \tag{9}$$

The expression on the right-hand side of the first equation is the aforementioned alternative index of kurtosis. It is strictly decreasing in b over a sufficiently wide range, hence the equation has a unique root $\hat{b}$.[11] On the other hand, the left-hand side of the equation shows that in the determination of $\hat{b}$ only the second, and no longer the fourth moment is involved, which confirms the expectation articulated above of more robustness.

The simple rule for rejection in (9) is a one-sided test (we are not interested in discriminating $b=2$ against high values of b). The critical values are readily obtained from 10,000 samples of T random draws from the standard normal. Computing the estimate $\hat{b}$ for each sample, it only remains to determine the 5% quantile of this collection. The specific values of T referred to will be the typical sample sizes in our empirical analysis.

[10] These estimations are not always without problems: for sample sizes less than 100 it may happen that the likelihood function has no minimum within a reasonable range (Agró, 1995, p. 527).

[11] Estimation of b *via* the implicit equation in (9) is also mentioned in Botazzi (1994, p. 4), the manual on the software SUBBOTOOLS, which is online freely available at `http://cafim.sssup.it/~giulio/software/subbotools`. It has, however, to be noted that there are two misprints in his formula (the software source code in file `subbofit.c`, on the other hand, is correct).

It should finally be mentioned that Fagiolo et al. (2008) also experimented with the Cauchy, the Student-t and the Lévy-Stable distribution as alternatives to fit fat- and medium-tailed distributions of output growth data. They conclude, however, that the EP density seems to outperform the other three density families.

3. Asymptotic results for the empirical data

Our empirical study is concerned with quarterly output data of the US economy, both in levels and first-differences. Regarding the former, we work with the CBO output gap, while for the growth rates we do not only refer to real GDP (in chained 2005 dollars) but also to the output of the firm sector, which for many models appears to be the more appropriate output concept.[12] The growth rates are annualized and denoted by 'gGDP' and 'gYF', respectively, the gap series is referenced as 'Gap'.

The data covers a period of not quite 50 years. It disregards the 1950s and begins in 1960:Q1, and we let it end in 2007:Q2 before the first signs of the financial crisis in the real sector. This amounts to a total of $T = 190$ quarters. Already on the basis of general observations, a part of macroeconomic research divides such an interval into two subsamples, which are commonly called the periods of the Great Inflation (GI) and the Great Moderation (GM). Estimations are then interested in possible parameter shifts in structural models, representing, for example, a different conduct of monetary policy in the two subperiods. With respect to output there is also rigorous econometric work showing a significant decline in its volatility. In fact, with quarterly growth rates from 1953:Q2 up to 1992:Q2, McConnel and Perez-Quiros (2000) reveal strong evidence for one—and only one—structural break, where the most suitable point estimate for the break date is 1984:Q1. We follow the upshot of their analysis and, maintaining the expressions GI and GM, distinguish two subperiods of almost equal length, GI: 1960:Q1 – 1983:Q4 ($T = 96$ observations) and GM: 1984:Q1 – 2007:Q2 ($T = 94$). The acronym for the full sample period is GIGM.

The main part of Table 1 computes the test statistics from the previous section for the three sample periods and the three original time series,. They are recorded in normal font size. If we first consider the full sample period then, as emphasized by the bold face figures, it leaps to the eye that the five statistics lead to very different conclusions concerning non-normality (also 'NN' henceforth). One statistic ($\hat{b}$) recognizes NN for all three series, and one (BN) rules it out for them. JB and AD conclude NN in two but not identical cases, GJB concludes it in only one case. So not only across the different types of test statistics but even within the class based on skewness and kurtosis (JB, GJB, BN), the evidence is rather mixed. The different outcomes illustrate that one should

[12] In essence, the 'firm sector' is non-financial corporate business. The data sources are given in the appendix.

	GIGM			GI			GM		
	Gap	gGDP	gYF	Gap	gGDP	gYF	Gap	gGDP	gYF
JB :	**7.69**	**11.86**	3.42	1.44	0.17	0.62	0.77	1.84	0.58
	24.73	**27.91**	**6.87**						
GJB :	1.66	**11.53**	3.28	0.38	0.17	0.61	0.17	1.73	0.55
	24.51	**27.64**	**6.80**						
BN :	1.44	4.09	1.39	0.80	0.26	0.56	0.59	2.72	0.46
	5.61	3.43	1.49						
AD :	0.68	**1.47**	**0.82**	0.29	0.36	0.37	0.22	0.62	0.60
	1.80	**1.19**	0.39						
$\hat{b}$:	**1.45**	**1.18**	**1.43**	2.07	1.91	1.84	2.54	**1.29**	**1.40**
	1.09	**1.25**	**1.57**	1.56	1.90	1.93	1.56	1.95	1.68

Table 1: Test statistics (empirical series and residuals from AR(p) estimations).

Note: Bold face figures indicate candidates for a rejection of normality. Normal fonts refer to the empirical series, small-sized fonts to the residuals from an AR(3) estimation (in case of the CBO gap) or an AR(2) estimation (in case of the growth rates gGDP and gYF). Regarding the first four test statistics, the latter entries are omitted in GI and GM since they offer no clue to non-normality.

be cautious with claims of "non-normality" that are referring to no more than one test statistic.

The results can be checked by purging the time series of their autocorrelation structure. Here the data are conceived of as being generated by a convenient stochastic process. General econometric studies often content themselves with AR(1) processes for this purpose, while Christiano (2007) was more ambitious and employed a four-lag VAR with seven variables (for monthly data, though). We believe that in the present context the uncorrelated residuals from AR(p) estimations are good enough, where for all three sample periods a lag length $p=3$ is sufficient for the CBO output gap, and $p=2$ for the two growth rates (parsimony in the number of parameters is not important in this respect, but higher lags yield no further noteworthy improvement in the fit).

The test statistics for these residuals are the entries in smaller font size in Table 1. Regarding GIGM both can happen, that normality is accepted for the depurated series and not for the original data (AD for gYF), and the other way around (in several cases). Most remarkable are GJB and AD for the output gap, according to which the innovations in the AR(3) process are strongly non-normal, while despite the linear structure this property does not carry over to the data produced by these shocks.

Almost any evidence of non-normality disappears when the shorter subperiods of GI and GM are considered, which holds for the original and the AR-filtered series alike. For all but the $\hat{b}$-test, the statistics are far from their critical values. Only for the two growth rates in GM (but not GI), values of $\hat{b}$ are obtained that fall short of the critical values given in eq. (9). The low value of $\hat{b} = 1.29$ for gGDP is especially striking since the residuals from the AR(2) estimation are almost perfectly normal according to this criterion.

4. Small-sample results for the empirical data

The critical value for $\hat{b}$ in (9), which in Table 1 lead us to a rejection of normality for gGDP and gYF in GM, was established under the null hypothesis of independent draws from a normal distribution. Since also for the uncorrelated AR(2)-residuals of the series one fails to reject normality of the shape parameter, doubts may arise as to whether the NN conclusion could be maintained once the serial correlation in the original series is properly taken into account (even if it is not overly strong). Regarding the conclusions from the four other test statistics, they are based on asymptotic theory and one may ask for their small-sample properties. These questions bring us to the second stage of the analysis, which is a battery of straightforward Monte Carlo (MC) experiments.

To this end we take the AR(p) estimates of a time series and simulate this process over the empirical sample size (after discarding a longer period at the beginning to rule out any transient effects). The innovations are independently drawn from the normal distribution with a variance equal to the estimation's squared standard error. This is repeated 10,000 times and for each of these MC samples the test statistics are computed.

By construction, the statistics should diagnose normality. According to the asymptotic theory for JB, GJB, BN, AD and the iid assumption for $\hat{b}$ in (9), the rules for rejecting normality should be false in just 5% of the 10,000 MC samples. Specifically, we have here stylized but empirically relevant conditions on time series data for which we can check how reliably this is done. In econometric terms, we can determine the so-called *size* of the five tests, that is, the probability of committing a type I error by rejecting normality when in fact this null hypothesis is true. This is the first kind of results presented in Table 2, which are independent of the values for the empirical test statistics in Table 1.

Related to this information are the quantiles of the collection of the simulated statistics; the 95% quantiles for JB, GJB, BN, AD and the 5% quantile for $\hat{b}$. These critical values for rejection in small samples under the present circumstances are reported in column *crit* in Table 2. The statistics *emp* of the empirical samples (reproduced from Table 1) can now be directly compared to them for a definite conclusion. In addition, we can compute what quantile q a value of *emp* constitutes in the MC distribution and obtain the p-value of the corresponding test statistic from it, which has the following interpretation:

if instead of *crit*, the value *emp* of the empirical statistic were employed as a benchmark for rejection, then for JB, GJB, BN, AD the percentage $p = 1-q$ would be the error rate of falsely rejecting normality, and for $\hat{b}$ it would be $p=q$. Certainly, at this paper's significance level non-normality would only be concluded with a p-value of less than 5%, and p-values above 5% would give us an idea of how safe we can feel when accepting normality.

	size	***crit***	***emp***	***p***	***size***	***crit***	***emp***	***p***
GIGM	Gap							
JB :	33.6	17.71	7.69	22.7				
GJB :	2.7	4.30	1.66	25.2				
BN :	5.5	6.09	1.44	72.6				
AD :	53.3	2.44	0.68	61.5				
$\hat{b}$:	12.1	1.39	1.45	6.7				
GIGM	gGDP				gYF			
JB :	4.6	5.78	11.86	**1.0**	4.8	5.84	3.42	14.4
GJB :	4.3	5.61	11.53	**0.9**	4.4	5.60	3.28	14.2
BN :	6.3	6.37	4.09	17.9	6.1	6.28	1.39	59.4
AD :	4.6	0.78	1.47	**0.1**	4.7	0.78	0.82	**3.9**
$\hat{b}$:	4.9	1.58	1.18	**0.0**	5.0	1.58	1.43	**1.1**
GM	gGDP				gYF			
JB :	4.3	5.58	1.84	32.8	4.4	5.57	0.58	72.7
GJB :	3.8	5.20	1.73	32.2	4.2	5.46	0.56	72.8
BN :	2.6	5.29	2.72	31.9	2.6	5.34	0.46	86.7
AD :	5.4	0.80	0.62	13.0	5.2	0.80	0.60	13.8
$\hat{b}$:	4.9	1.46	1.29	**1.0**	5.1	1.45	1.40	**3.2**

Table 2: Statistics from AR(p) simulations with normal innovations.

Note: Based on 10,000 MC samples for each series and subperiod. Size and p-values in per cent, column '*crit*' indicates the critical quantiles of the MC distributions (95% and 5%, respectively), and '*emp*' the empirical statistics from Table 1.

Table 2 is limited to the series and sample periods for which non-normality was not outright denied by the tests in Table 1. To begin with the evaluation of the size in Table 2, its values for the single statistics (except perhaps BN) are all satisfactorily close to the nominal level of 5% if we look at the two growth rates. By contrast, there are dramatic deviations for JB and AD when these tests are applied to the CBO output gap. Regarding JB this is due to an increase of the asymptotic rejection level from $\chi^2(2) = 5.99$ to a 95% quantile of 17.71, for AD the previous standard level of 0.787 increases here to 2.435. The decline of the 5% level of $\hat{b}$ from 1.578 (for $T = 190$) to 1.394, which raises the size of this test to 12.1%, is a weaker but still unpalatable phenomenon. These deteriorations are mainly caused by the strong serial correlation of 0.93 in the output gap, which of course are accounted for by the AR(p) coefficients in the simulations. In comparison, the size effects from the correlation between 0.23 and 0.30 in the growth rates appear rather unimportant.

The approach of capturing the autocorrelation structure of the output gap by the Monte Carlo simulations has also consequences for the NN conclusions. Now there is no statistic left that would recommend a rejection of normality; only $\hat{b}$ has a p-value that is above but not too far away from the 5% level. Therefore, to sum up our evidence concerning the CBO output gap, we do not have sufficient support of any non-normal features in this series.[13]

Prospects of non-normality are better for the growth rates of output. Regarding the full sample period, four of the five p-values reject normality for gGDP (even strongly so), and normality of the firm sector growth rates is rejected by AD and $\hat{b}$. Regarding the GM subsample, recall that only the $\hat{b}$ test provided evidence against normality in Table 1. This tentative result is now fully confirmed by the Monte Carlo experiment; cf. the bottom part of Table 2. The fact that here the three statistics based on skewness and kurtosis do not offer the least clue against normality suggests that the test based on the EP distributions measures something more general. Given the p-values of 13.0% and 13% for AD, this test, which likewise considers the shape of an entire distribution, seems to be somewhere in the middle between the two principles. Hence summarizing claims of non-normal growth rates can only be properly assessed with additional information about the particular specification of their non-normality.

Finally, we can also try to make sense of the fact that for gGDP in GM the estimated $\hat{b} = 1.29$ reveals a non-negligible non-normality, whereas the estimated innovations are nearly normally distributed ($\hat{b} = 1.95$ in Table 1). With the linear AR(2) filter, as we have just seen for $b = 2$, this should be an extremely rare event. A tentative alternative conclusion could thus be that generally an AR(p) process, or even a multivariate VAR,

[13] This finding, in particular, means that De Grauwe's (2012) reference to a non-normal output gap, which he points out his model is successfully able to reproduce, is not warranted; not even if one goes beyond his argumentation with the kurtosis or the Jarque-Bera statistic. So far, however, this observation only indicates that the issue of "non-normality" requires further discussion of what his model should more precisely achieve.

may not be a good hypothesis, and that the strong differences in the two statistics are rather indicative of a strong nonlinear mechanism in this sample period.

5. A two-regime Monte Carlo experiment

As already pointed out by McConnel and Perez-Quiros (2000), the unique structural break they identify consists in a significant decline in output volatility. This phenomenon is also clearly visible in our three empirical series, even in time series diagrams. Table 3 documents a decrease of about one-half from GI to GM not only in the standard deviations of the empirical data, but also in the standard deviations of the residuals from the AR(p) estimations. One needs no formal statistical tools to classify this as a significant change. On the other hand, the changes in the mean growth rates or the AR(p) slope coefficients are much more moderate.

Gap		gGDP		gYF	
GI	GM	GI	GM	GI	GM
2.93	1.43	4.45	2.12	4.12	2.62
0.98	0.44	4.28	1.90	3.89	2.47

Table 3: Standard deviations of the empirical series and their AR(p) residuals (upper and lower row, respectively).

Intuitively, if we have a time series with strong noise in the first half and weak noise in the second half, one may suppose that this gives rise to a higher kurtosis, since the higher values (in modulus) in the first half become a rarer event when considering the full sample. The argument should be valid even if the random forces are normally distributed in each of the two subsamples, and a certain amount of non-normality may also be indicated by other test statistics. The conjecture that a structural change in the volatility may contribute to the non-normality results above can, however, be readily checked by another Monte Carlo experiment.

Consider the GDP growth rates over GIGM and take the NN results based on JB, GJB and AD in Table 2 as an example. The simulated data generation process was an AR(2) over the full length of $T = 190$ quarters, where the normal innovations yielded values of the test statistics that are in their vast majority below the empirical estimates (conversely for the $\hat{b}$ statistic). We now maintain the null hypothesis of normal innovations but introduce the regime shift from GI to GM into the simulations. That is, the first $T_{GI} = 96$ periods of the simulation (again after a suitable transitory period) adopt the AR(2) coefficients estimated over GI, and the second $T_{GM} = 94$ periods use the coefficients as they have been estimated over GM. This, in particular, includes the different variances for the normal random draws (which according to Table 3 are equal to $(4.28)^2$ and $(1.90)^2$, respectively).

	Gap			gGDP			gYF		
	crit	*emp*	*p*	*crit*	*emp*	*p*	*crit*	*emp*	*p*
JB :	65.48	7.69	49.9	50.49	11.86	45.5	22.25	3.42	46.9
GJB :	14.92	1.66	51.9	48.45	11.53	44.9	20.94	3.28	46.5
BN :	7.01	1.44	75.7	8.12	4.09	47.1	5.90	1.39	68.9
AD :	4.66	0.68	81.6	2.27	1.47	25.4	1.21	0.82	17.9
$\hat{b}$:	1.02	1.45	38.9	1.09	1.18	14.8	1.32	1.43	14.8

Table 4: Two-regime Monte Carlo simulations, pooling AR(p) from GI and GM.

The question we then ask is whether or to what extent the structural break tends to increase the simulated JB, GJB, AD statistics (and to decrease the values of $\hat{b}$). The answer in the middle of Table 4 is unambiguous. Compared to the experiment in Table 2, we observe a drastic increase in the critical 95% quantiles of these statistics (and a sizeable decrease of the 5% quantile of $\hat{b}$). The quantiles even exceed the corresponding empirical value (or fall short of it in case of $\hat{b}$), with the consequence that the p-values distinctly rise above the 5% levels.

The same effects are obtained for the growth rate gYF and, for completeness, the output gap. Overall, there is no single p-value lower than even 10%. It can therefore be said that the previous findings of non-normality over the full sample period may be spurious; all evidence of non-normality could be explained by normal innovations in a simple linear stochastic process once one takes account of the significant decline in their volatility. While admittedly, the experimental design with the sudden change at the break data is somewhat crude, the high p-values in Table 4 do not give us much reason to expect that a smoother transition from one regime to the other would lead to an essential weakening of the conclusion.

6. The Laplacian as an alternative hypothesis

Because of the unconvincing evidence of non-normality over GIGM and since in the subsamples GI and GM the only indicator of a possible non-normality is the shape parameter b of the EP distribution, we will in the remainder of the paper be exclusively concerned with this statistic. Given that for the growth rates in GM normality was strongly rejected by the tests in Table 2, we may put forward an alternative hypothesis. As mentioned above, already for systematic reasons the Laplace distribution with $b=1$ is usually considered to be a natural antithesis to the normal distribution with $b=2$. As furthermore the growth rate estimates $\hat{b} = 1.29$ and $\hat{b} = 1.40$ are closer to 1 than to 2, $b=1$ does not appear to constitute an unreasonable alternative benchmark distribution.

Let us begin with a geometric account of the goodness-of-fit of the shape parameter estimations for the growth rates in GI and GM. More specifically, we can also gain an intuitive impression of how far their distributions are from the two benchmark distributions. To this end, we draw the densities functions of the standardized GDP growth rates z in the semi-log diagrams in Figure 1, where with respect to the estimated scale and location parameters $\hat{a}$ and $\hat{m}$ and the original growth rates $x = \text{gGDP}$, the standardized values are given by $z = (x - \hat{m})/\hat{a}$.[14] The diagrams for the firm sector growth rates do not look very different, so Figure 1 is sufficiently representative.

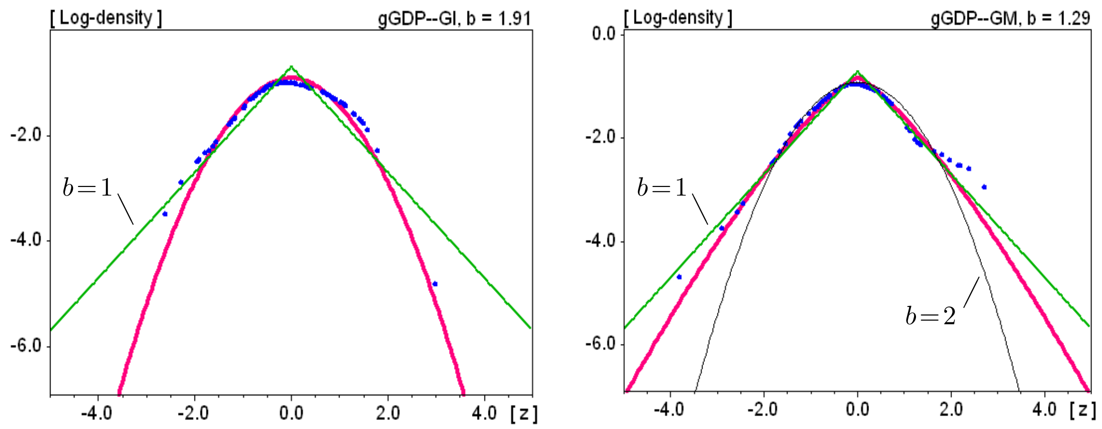

Figure 1: Estimated densities (EP/nonparametric densities: solid lines/dots).

The left-hand panel in Figure 1 deals with the GI subsample. The exponential power density constituted by the estimated shape parameter $\hat{b} = 1.91$ is drawn as the bold (red) line. The dots distributed around it are the density values of the $T = 96$ observations of this period. That is, we use a standard nonparametric approach to compute a kernel density estimator $\hat{f} = \hat{f}(z)$ of the empirical values of z and for each observed z_t plot the point $(z_t, \hat{f}(z_t))$.[15] Over a wide range and especially in the middle part, these points

[14] A prescription of how $\hat{a}$ and $\hat{m}$ are obtained after $\hat{b}$ has been estimated before can be found in the Appendix.
[15] We employ the Epanechnikov kernel for this purpose; see Davidson and MacKinnon (2004, pp. 678–683) for the computational details.

really nestle into the smooth curve of the theoretical density function. The estimation thus inspires confidence, its outcome being more than just the result of a somewhat abstract and technical concept.

The log-density of the normal distribution looks very similar to the EP density with $b = 1.91$, which is the reason why it has been omitted in this diagram. The other polar case of the Laplacian density, where $b=1$, is the tent-shaped thin solid line. Over most of the empirical range it yields a clearly worse fit. One might nevertheless argue that it could perhaps provide better results for the moderate and more extreme negative values of z_t, which overall may even suggest an asymmetric estimation approach (as it was hinted at in fn 9) with different values of b for positive and negative values of z. Here, we mainly abstain from this idea since given the relatively small sample size, this would also require an econometric discussion of the risk of overfitting.

The right-hand panel in Figure 1 presents the results for the GM period. The estimated density function with $\hat{b} = 1.29$ spreads wider away from that of the normal distribution (the curve with $b=2$ in the diagram), thus giving greater weight to the outer regions. To some part this seems to be an implication of the sharper turnaround in the centre; to another part the shallower slope of the points $(z_t, \hat{f}(z_t))$ with z_t between 1 and 3 may contribute to it, although there are again no observations with $|z_t| \geq 4$. Both the outer positive and negative points in the diagram indicate that now even the Laplacian density would not accomplish too bad a fit.

The latter observation prompts the idea of investigating the Laplacian as a specific alternative hypothesis to normality. More precisely, we can put forward $b=1$ as another null hypothesis and test it with similar Monte Carlo experiments to Table 2, now with re-estimations of the $\hat{b}$-statistic only. However, we wish to weaken the underlying structural assumptions, that is, we no longer estimate an AR(p) process (now with Laplacian innovations). Instead, we randomly draw directly from the standardized Laplace distribution, though in such a way that also the empirical first-order autocorrelation ρ is taken into account. [16] Of course, the normal distribution ($b=2$) can be dealt with in the same way.

We run this experiment with the sample sizes $T = 96$ and $T = 94$ for GI and GM, respectively, again 10,000 simulation runs for each case. [17] Re-estimating the shape parameter $\hat{b} = \hat{b}^c$ for each such run c ($c = 1, \ldots, 10,000$), a distribution $\{\hat{b}^c\}$ is obtained to which we can relate the empirical estimate $\hat{b}$ from above. The tests that we thus carry out are one-sided. We reject normality if this $\hat{b}$ is below the 5% quantile $Q_{0.05}$ of the MC

[16] This sampling could be viewed as an AR(1) process with Laplacian innovations, the variance of which is linked to ρ (and, however, ρ only); see the Appendix for this and the generation of iid. pseudo random numbers from an EP distribution. Considering the standardized Laplace distribution suffices since the estimation of b is independent of the other parameters m and a.

[17] Apart from the strong tendency towards normality documented in Table 1, subjecting the CBO output to the same exercise is not very informative because its high autocorrelation of $\rho = 0.93$ leads to an extremely wide dispersion in the re-estimated values of b.

	empirical		null: $b = 2$		null: $b = 1$	
	ρ	$\hat{b}$	$Q_{0.05}$	p	$Q_{0.95}$	p
gGDP						
GI :	0.26	1.91	1.44	**36.8**	1.68	1.5
GM :	0.23	1.29	1.44	1.6	1.65	**29.1**
gYF						
GI :	0.30	1.84	1.44	**31.0**	1.73	2.9
GM :	0.26	1.40	1.43	3.9	1.69	**20.0**

Table 5: Testing the normal ($b\!=\!2$) and the Laplace ($b\!=\!1$) distribution.

Note: Each case based on 10,000 Monte Carlo samples of T random draws from the EP distribution with shape $b\!=\!1$ or $b\!=\!2$, respectively, correlated with coefficient ρ (T being the empirical sample size). $Q_{0.05}$ and $Q_{0.95}$ are the 5% and 95% quantiles of the MC distribution of the re-estimated $\hat{b}$. p-values in per cent and for one-sided tests, that is, for $b\!=\!2$ ($b\!=\!1$) p is the percentage of values in the MC distribution that are less than the empirical $\hat{b}$ (larger than this $\hat{b}$).

distribution under the null of $b\!=\!2$, and we reject the Laplacian if this $\hat{b}$ is above the 95% quantile $Q_{0.95}$ of the distribution under the null of $b = 1$. The p-values are determined accordingly.

The results presented in Table 5 leave us a clear and pronounced message. For both growth rate series alike, as emphasized by the bold face figures, it can be concluded that over the GI period the Laplacian is rejected and normality is accepted. By contrast, over GM it is just the other way around: here normality is rejected and the Laplacian is accepted, in the sense that it cannot be ruled out.

These results are based on the particular values of the autocorrelation in the empirical data. Within the typical range of serial correlation in the quarterly growth rates of aggregate output, however, the critical quantiles of the MC distributions $\{\hat{b}^c\}$ remain quite insensitive. Regarding the choice between the normal and the Laplace distribution we can therefore conclude this subsection with putting forward a simple rule of thumb. Referring to a correlation coefficient $\rho = 0.25$ and a sample size $T\!=\!95$, it reads,

$$\begin{array}{lll} \hat{b} > 1.671 & \text{reject } b = 1 & \\ \hat{b} < 1.438 & \text{reject } b = 2 & \\ \hat{b} \in [1.438, 1.671] & \text{compatible with both } b\!=\!1 \text{ and } b\!=\!2 & \end{array} \qquad (10)$$

Perhaps easier to recall, we can also say that the inconclusive range is given by 1.55±0.12, while above that interval normality may be accepted and below it a Laplace distribution.

7. On the precision of the estimates of the shape parameter

Regarding the stylized facts of the macro economy, a model may be judged by, *inter alia*, how well its output growth rates are able to reproduce the empirical shape of the EP distribution. In order to put the model's degree of matching into perspective, we need to know something about the precision of the estimation of $\hat{b}$. Conventionally, we thus ask for the standard error of $\hat{b}$. Readily available for this is the asymptotic variance, which can be explicitly computed as

$$\mathrm{Var}(\hat{b}) \;=\; \frac{\hat{b}^3}{(1+1/\hat{b})\,\Psi'(1+1/\hat{b}) \;-\; 1} \tag{11}$$

where $\Psi'(\cdot)$ is the trigamma function, i.e. the second derivative of the logarithm of the Gamma function (Agró, 1995, pp. 524f; Bottazzi and Secchi, 2008, p. 5). As it should be, $\mathrm{Var}(\hat{b})$ is independent of the location and scale of the distribution. On the other hand, it changes with the level of the estimate. While the denominator in (11) is rising with $\hat{b}$, the increase in the numerator is stronger. Hence the more normal the distribution, so to speak, the higher the variance. These variations are sizeable. For example, $\hat{b}=1$ and $\hat{b}=2$ give rise to a variance of 3.45 and 19.89, respectively, meaning that the standard error more than doubles.

Now, one may be sceptical about employing (11), not only because of the relatively small size of our samples but also since it derives from the maximum likelihood estimation of independent random draws from the EP distribution. In addition to (11), we therefore make use of two bootstrap procedures to determine the confidence intervals around $\hat{b}$.

A first and obvious approach takes up the Monte Carlo experiments in the previous section where 10,000 samples of autocorrelated data were generated under the null hypothesis of $b = 1$ and $b = 2$. Here we only have to replace these polar values with the empirical estimates $\hat{b}$. This procedure can be viewed as a parametric bootstrap. The standard deviation of the collection of the re-estimated values $\{\hat{b}^c\}$ gives us the bootstrapped standard error, and suitable quantiles of it the lower-and upper-bounds of a confidence interval.

The second approach is a nonparametric bootstrap that directly samples from the empirical data set with its T observations. Because of the serial correlation, three block bootstraps BB1, BB4, BB10 are considered which sample (with replacement, of course) from the overlapping blocks of length 1 (the degenerate case), 4 and 10, respectively. These frequency distributions are likewise referred to as $\{\hat{b}^c\}_{c=1}^{10,000}$.

Table 6 reports some basic indicators of the dispersion of the bootstrap distributions, namely, the 5% and 95% quantiles and the standard deviation, i.e. the bootstrapped

	gGDP				gYF			
	emp	$Q_{0.05}$	$Q_{0.95}$	*ser*	*emp*	$Q_{0.05}$	$Q_{0.95}$	*ser*
GI	1.91				1.84			
PB :		1.40	3.29	0.633		1.37	3.19	0.608
BB1 :		1.36	3.17	0.600		1.34	3.12	0.590
BB4 :		1.35	3.18	0.609		1.38	2.96	0.525
BB10 :		1.39	3.27	0.627		1.38	2.96	0.524
$\sigma_b^\star$:				**0.442**				**0.395**
GM	1.29				1.40			
PB :		1.03	2.12	0.349		1.12	2.38	0.409
BB1 :		1.05	1.88	0.282		1.08	2.10	0.341
BB4 :		1.08	1.84	0.240		1.10	2.11	0.332
BB10 :		1.10	1.78	0.215		1.09	2.28	0.383
$\sigma_b^\star$:				**0.189**				**0.293**

Table 6: Dispersion statistics of the bootstrapped distributions $\{\hat{b}^c\}$.

Note: Sample size of the $\{\hat{b}^c\}$: 10,000. $Q_{0.05}$ and $Q_{0.55}$ are their 5% and 95% quantiles, *ser* the standard error, and *emp* the empirical estimate. PB is the parametric bootstrap with the empirical autocorrelation, BBn the block bootstrap with block length n ($n = 1, 4, 10$), while $\sigma_b^\star$ is the standard deviation that equates the normal density at its mode to the (estimated) density of BB10 at its mode.

standard error of the estimation. Once again, these statistics are computed for GI and GM, and for the output growth rates of GDP and the firm sector. What holds for all four cases is that the confidence intervals are not symmetric around the estimated value of b: the positive deviations of $\hat{b}^c$ from the estimate $\hat{b}$ are larger than the negative deviations. These distortions can be even so serious that the standard errors would be misleadingly high. In the calculation of the latter we have therefore truncated the $\hat{b}^c$ at 5.

Concerning the question of whether the data could be compatible with the Laplace distribution, the answer from the confidence intervals is strongly negative for the GI period and weakly negative for GM. The normal distribution can be dismissed for the GDP growth rates in GM if we consult the block bootstraps, but not if the parametric bootstrap is employed. For gYF, none of the bootstrap distributions stays away from $b = 2$. Generally, owing to the short sample periods, the precision of the estimations is so limited that we have to be satisfied if at least one of the polar cases $b = 1$ or $b = 2$ is ruled out. However, this may contradict the results from Table 5, which for GM could not reject the hypothesis $b=1$ whereas here the 5% quantiles of $\{\hat{b}^c\}$ are bounded away from unity.

Because of the asymmetry already noted, the standard error has to be interpreted with caution. On the other hand, such a statistic is not only a succinct information but also of practical use. In assessing the deviations of a model-generated $\hat{b}$ from its empirical counterpart and making them comparable to the deviations of other summary statistics (also called moments), these magnitudes have to be suitably weighted; in particular, if, following Ruge-Murcia (2012) mentioned in the Introduction, a model is estimated by the method of simulated moments. The common treatment is here a multiplication of the model deviation with the inverse of the variance of the moment. This controls for the scale and the weight is proportionately higher the higher the precision of the estimate, or the lower its variance.

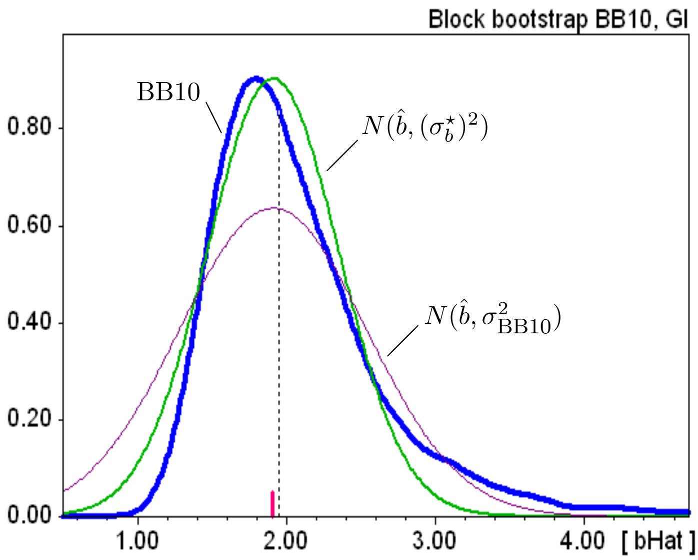

Figure 2: Bootstrapped density BB10 of the shape parameter estimate $\hat{b}$ ('bHat').

Note: Bootstrap of the GDP growth rates in GI. The dotted line indicates the median of BB10, the bold (red) bar at the bottom the empirical estimate $\hat{b}$.

A problem arises with asymmetric distributions like our bootstraps since then the variance may imply a certain overstatement, such that the weight of this moment is unduly low. To discuss this issue, consider the bold line of the distribution BB10 for gGDP over GI in Figure 2 (BB10 is quite as good as any of the other bootstrap distributions). The kind of asymmetry is clarified by the vertical dotted line, which is the median of the distribution. The idea of the conventional weighting factor rests on the supposition that the corresponding moment is nearly normally distributed around $\hat{b}$, at least in some neighbourhood of this estimate. If we try to approximate such a distribution by a normal distribution around $\hat{b}$ with the variance σ^2_{BB10} of the bootstrap, we see that it assigns probabilities that are far too low for values near $\hat{b}$, and too high for estimates less than 1.30 (roughly).

Clearly, responsible for these distortions is too low a value of this normal density function at $\hat{b}$. A better approximation would be a normal distribution that is equally high in the centre as BB10 at its mode. Accordingly, we specify $\sigma^\star_b$ as the value that renders the density of $N(\hat{b}, (\sigma^\star_b)^2)$ at $\hat{b}$ equal to the value of the estimated density function of BB10 at its mode. Figure 2 illustrates that over the main range of BB10, this normal distribution is indeed a suitable approximation. Hence, if reference is made to a single standard error, this mode-consistent standard error $\sigma^\star_b$ appears to be the most appropriate concept. Its values are shown as the boldface figures in Table 6, certainly all of them being smaller than the standard deviations of the asymmetric bootstrap distributions.

$\hat{b}$:	1.00	1.29	1.40	1.84	1.91	2.00
$\sigma^\star_b$:	—	0.189	0.293	0.395	0.442	—
$\sqrt{\text{Var}(\hat{b})\,/\,T}$:	0.191	0.262	0.291	0.408	0.429	0.458

Table 7: Mode-consistent and asymptotic standard errors for $\hat{b}$.

Note: Underlying are $T = 96$ for $\hat{b} = 1.84, 1.91$ (GI), $T = 94$ for $\hat{b} = 1.29, 1.40$ (GM), and $T = 95$ for the benchmark cases $\hat{b} = 1, 2$.

Even if the specification of $\sigma^\star_b$ is not backed up by rigorous econometric theory, we believe that the intuitive argument of Figure 2 makes good sense. To put the results obtained for the mode-consistent standard errors into perspective, we should nevertheless return to the asymptotic variance $\text{Var}(\hat{b})$ in (11) and compare the corresponding errors $[\text{Var}(\hat{b})/T]^{1/2}$ to them. This is done in Table 7. It shows that three of the four $\sigma^\star_b$ are amazingly close to the asymptotic standard errors. The difference between the two statistics is larger for gGDP in GI, where $\hat{b} = 1.29$, but the ranking across the $\hat{b}$ is still preserved. On the whole, the table suggests that despite the small samples and the neglect of serial correlation, the asymptotic standard error is not so unreliable after all. At least as far as the information content of a single statistical number is concerned, one may thus save the effort of a bootstrapping procedure and invoke (11) directly.

8. Conclusion

The primary motivation for this paper were recent claims in the literature that US output growth rates fail to be normally distributed. However, although all of these studies have long sample periods of more than 40 years in common, none of them mentions the possibility of a structural break that might or might not invalidate the results. A straightforward Monte Carlo experiment could demonstrate that this neglect is indeed unwarranted. Simulating two AR(p) processes for the two periods of the Great Inflation (GI) and Great Moderation (GM) with their different variances of the—normally distributed—innovations, the pasted growth rate series is typically found to exhibit non-normal behaviour of the type measured in the empirical data. Therefore, as long as no new and more sophisticated evidence is provided, the previous results of non-normality appear to be spurious.

Correspondingly, we were then looking for non-normality within the two shorter subsamples. Of the five test statistics that we considered, the only indication of non-normality is the estimated shape parameter $\hat{b}$ of the exponential power distribution over GM, which distinctly falls short of the benchmark for normality, $b=2$. Here it has to be noticed that this finding did not take the serial correlation in the data into account, and that the estimated $\hat{b}$ for the residuals of a suitable AR(p) process does not essentially deviate from $b=2$, or only moderately so. This suggests that if it is normally distributed innovations that ultimately drive the economy, the transmission mechanisms may be of a distinctly nonlinear nature. Alternatively, of course, there may also be non-normalities in the shock processes themselves.

As a first and largely atheoretical step in this direction, we put forward the hypothesis that the growth rates were obtained from a normal ($b = 2$) *vis-à-vis* a Laplace ($b = 1$) distribution with its fatter tails, where both of them exhibit the empirical autocorrelation. The p-values from this Monte Carlo experiment allowed us the interpretation that normality prevailed in GI, whereas the Laplace distribution took over in GM.

This is a nice and pronounced statement that has not been put up to discussion before. On the other hand, it has to be admitted that the message cannot be fully maintained from the perspective of the bootstrapped confidence intervals around the estimated $\hat{b}$. This qualification may be taken as a final example that any claim of non-normality, for longer or shorter sample periods, requires an additional discussion of the specific measurement approach.

Appendix

Data sources

The data of potential output are from the report "The Budget and Economic Outlook: Fiscal Years 2012–2022" (January 2012), downloadable at `http://www.cbo.gov/publication/42912`. Real GDP was obtained from the Bureau of Economic Analysis, at `http://www.bea.gov/national/index.htm#gdp`.

The firm sector output series was extracted from the database `fmdata.dat` in the zip file `fmfp.zip`, provided by Ray Fair for working with his macroeconometric model. It is a plain textfile downloadable from `http://fairmodel.econ.yale.edu/fp/fp.htm`. The acronym to identify the series is 'Y', as explained in Appendix A.4, Table A.2., of the script *Estimating How The Macroeconomy Works* by R.C. Fair, January 2004, which can be downloaded from `http://fairmodel.econ.yale.edu/rayfair/pdf/2003a.pdf`.

Approximation of the standard normal cumulative distribution function

Let $\phi = \phi(x)$ be the probability density function of the standard normal, $\phi(x) = \exp(-x^2/2)\,/\,\sqrt{2\pi}$, and $\Phi = \Phi(x)$ the standard normal cumulative distribution function. Then according to Abramowitz and Stegun (1964, p. 932, algorithm 26.2.17), up to an absolute error $|\varepsilon(x)| < 7.5 \cdot 10^{-8}$, the latter is approximated as follows,

$$\Phi(x) = \begin{cases} 1-c & \text{if } x \geq 0 \\ c & \text{if } x < 0 \end{cases} \qquad \text{where}$$

$$c = \phi(x)\,[\,b_1 z + b_2 z^2 + b_3 z^3 + b_4 z^4 + b_5 z^5\,] \;+\; \varepsilon(x)$$

$$z = z(x) = 1\,/\,(\,1 + b_0\,|x|\,)$$

$$\begin{aligned} b_0 &= 0.2316419 & b_1 &= 0.319381530 \\ b_2 &= -0.356563782 & b_3 &= 1.781477937 \\ b_4 &= -1.821255978 & b_5 &= 1.330274429 \end{aligned}$$

Estimation of *a* and *m* for an exponential power distribution

Suppose that the shape parameter b has already been estimated before as described by eq. (9) in the main text. Setting, in a ML estimation, the partial derivative of the log-likelihood function with respect to m equal to zero, $\hat{m}$ can then be obtained as the solution of the implicit equation in m,

$$\sum_{t=1}^{T} |x_t - m|^{\hat{b}-1} \;\; \text{sgn}(x_t - m) \;\;=\;\; 0$$

where sgn is the sign function, $\text{sgn}(y) = 1$ $(0, -1)$ if $y > 0$ $(y = 0$ or $y < 0$, respectively); cf. Mineo (2003, p. 112). On this basis, Chiodi (1988) has proposed the following expression as an unbiased estimate of a,

$$\hat{a} \;=\; \Big[\, \frac{\sum_{t=1}^{T} |x_t - \hat{m}|^{\hat{b}}}{T - \hat{b}/2} \,\Big]^{1/\hat{b}}$$

(quoted from Mineo and Ruggieri, 2005, p. 4, eq. (9)).

Random variates from EP distributions

In general, the generation of pseudo random numbers drawn from an EP distribution involves draws from a Gamma distribution, which in turn requires some computational effort (see, e.g., Zhu and Zinde-Walsh, 2009, p. 91, or Li, 2011, Section 2, which both allow for an asymmetric shape also). For the class of standardized distributions with shape $b > 1$ (besides $m = 0$, $a = 1$), Chiodi (1995, Section 4) set up a faster and easy-to-implement algorithms which has the advantage that it only needs the generation of uniformly distributed random numbers. A random number z is here generated in the following two stages: [18]

1. Repeat
   draw U and V from the uniform distribution over $[-1, +1]$
   and put $W = |U|^b \;+\; |V|^{b/(b-1)}$
   until $W \leq 1$.
2. Put $z = U \cdot [-b \ln(W)/W]^{1/b}$.

While the procedure fails to be applicable to $b = 1$, we checked that it is robust and works well for values of b arbitrarily close to unity. In our experiments with $b = 1$ it is thus perfectly sufficient to have recourse to the approximation $b = 1.00001$.

Of course, the draws thus obtained are iid. To take account of an autocorrelation ρ put, in round t, $z_t = \rho\, z_{t-1} + \sqrt{1 - \rho^2}\, \tilde{z}$, where besides $|\rho| < 1$ it is supposed that z_{t-1} is a draw from the previous round and $\tilde{z}$ a draw from the EP distribution, both of them with the same variance σ^2. It is easily seen that then $\mathrm{Var}(z_t) = \sigma^2$ and $\mathrm{Corr}(z_t, z_{t-1}) = \rho$. It is well-known that for normal distributions, $b = 2$, z_t is normally distributed, too. We know of no mathematical proof that establishes the analogous statement for general values of b. The property can, however, be confirmed by simulation studies, even for b close to one, although (very) large samples are required for a satisfactory convergence of the sample density function towards the theoretical density (the smaller b or the higher ρ, the larger the samples).

[18] The procedure can still be accelerated by suitable squeeze methods, at the price of a more complicated computer code. Since the original version is already fast enough, this does not seem worth the effort.

References

Abramowitz, M. and Stegun, I. (1964): *Handbook of Mathematical Functions.* Electronic copy of p. 932 at
`http://people.math.sfu.ca/~cbm/aands/page_932.htm` .

Agró, G. (1995): Maximum likelihood estimation for the exponential power function parameters. *Communications in Statistics—Simulation and Computation*, 24:2, 523–536.

Alfarano, S., Milaković, M., Irle, A. and Kauschke, J.: A statistical equilibrium of competitive firms. *Journal of Economic Dynamics and Control*, 36:1, 136–149.

Ascari, G., Fagiolo, G. and Roventini, A. (2012): Fat-tail distributions and business-cycle models. Working paper series, Department of Economics, University of Verona.

Bai, J. and Ng, S. (2005): Tests for skewness, kurtosis, and normality for time series data. *Journal of Business and Economic Statistics*, 23:1, 49–60.

Blatt, J.M. (1983): *Dynamic Economic Systems: A Post-Keynesian Approach.* Sharpe: Armonk, N.Y.

Bottazzi, G. (2004): Subbotools user's manual. LEM Working Paper Series 2004/14, Sant'Anna School of Advanced Studies, Pisa.

Botazzi, G. and Secchi, A. (2008): Maximum likelihood estimation of the symmetric and asymmetric exponential power distribution. LEM Working Paper Series 2008/19, Sant'Anna School of Advanced Studies, Pisa.

Chiodi, M. (1988): Sulle distribuzioni di campionamento delle stime di massima verosimiglianza dei parametri delle curve normali di ordine p. Technical report, Istituto di Statistica, Facoltà di Economia e Commercio di Palermo.

Chiodi, M. (1995): Generation of pseudo random variates from a normal distribution of order p. *Statistica Applicata (Italian Journal of Applied Statistics)*, 7:4, 401–416.

Christiano, L.J (2007): Comment. *Journal of Business and Economic Statistics*, 25:2, 143–151.

Davidson, R. and MacKinnon, J.G. (2004): *Econometric Theory and Methods.* Oxford University Press, Oxford.

De Grauwe, P. (2012): Booms and busts in economic activity. *Journal of Economic Behavior and Organization*, 83:3, 484–501.

DeLong, J.B. and Summers, L.H. (1986): Are business cycles symmetrical? In R.J. Gordon (ed), *The American Business Cycle: Continuity and Change.* University of Chicago Press, Chicago; pp. 166–179.

Fagiolo, G., Napoletano, M., Piazza, M. and Roventini, A. (2009): Detrending and the distributional properties of U.S. output time series. *Economics Bulletin*, 29:4, 3155–3161.

Fagiolo, G., Napoletano, M. and Roventini, A. (2008): Are output growth-rate distributions fat-tailed? Some evidence from OECD countries. *Journal of Applied Econometrics*, 23, 639–669.

Fagiolo, G., Napoletano, M. and Roventini, A. (2007): How do output growth-rate distributions look like? Some cross-country, time-series evidence. *The European Physical Journal B*, 57, 205–211.

Greene, W.H. (2002): *Econometric Analysis*, 5th edition. Prentice-Hall, Upper Saddle River, N.J.

Jarque, C.M. and Bera, A.K. (1980): Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6:3, 255–259.

Lobato, I.N. and Velasco, C. (2004): A simple test of normality for time series. *Econometric Theory*, 20:4, 671–689.

McConnell, M.M. and Perez-Quiros, G. (2000): Output fluctuations in the United States: What has changed since the early 1980's? *American Economic Review*, 90:5, 1464–1476.

Mineo, A.M. (1994): Un nuovo metodo di stima di p per una corretta valutazione dei parametri di intensità e di scala di una curva normale di ordine p. *Atti della XXXVII Riunione Scientifica della Società Italiana di Statistica*, CISU, Roma, pp. 147–154.

Mineo, A.M. (2003): On the estimation of the structure parameter of a normal distribution of order p. *Statistica*, 63:1, 109–122.

Mineo, A.M. and Ruggieri, M. (2005): A software tool for the exponential power distribution: The normalp package. *Journal of Statistical Software*, 12:4, 1–24.

Ruge-Murcia, F. (2012): Estimating nonlinear DSGE models by the simulated method of moments: With an application to business cycles. *Journal of Economic Dynamics and Control*, 36:6, 912–938.

Vavra, M. and Psaradakis, H. (2011): Testing normality in time series. *Mimeo*, Birkbeck College, University of London.

zhu, D. and Zinde-Walsh, V. (2009): Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics*, 148:1, 86–99.