

# ***Research Report***

***Department of Statistics***

**No. 2008:5**

**On the Probability Distribution  
of Economic Growth**

**Pär Stockhammar  
Lars-Erik Öller**

# On the Probability Distribution of Economic Growth

Pär Stockhammar and Lars-Erik Öller

Department of Statistics, Stockholm University  
S-106 91 Stockholm, Sweden  
E-mail: par.stockhammar@stat.su.se

## Abstract

Normality is often mechanically and without sufficient reason assumed in econometric models. In this paper three important and significantly heteroscedastic GDP series are studied. Heteroscedasticity is removed and the distributions of the filtered series are then compared to a Normal, a Normal-Mixture and Normal-Asymmetric Laplace (NAL) distributions. NAL represents a reduced and empirical form of the Aghion and Howitt (1992) model for economic growth, based on Schumpeter's idea of creative destruction. Statistical properties of the NAL distributions are provided and it is shown that NAL competes well with the alternatives.

Keywords: The Aghion-Howitt model, asymmetric innovations, mixed normal- asymmetric Laplace distribution, Kernel density estimation, Method of Moments estimation.

## 1. Introduction

In the Schumpeterian world growth is driven endogenously by investments into R&D, leading to better products, which initially capture monopoly profits. The quality improvements occur randomly over time. The main contributions to endogenous growth are Romer (1986) and Lucas (1988). They both argued that the underlying growth is determined by the accumulation of knowledge, with occasional setbacks. Other important papers on endogenous growth are: Segerstrom, Anant and Dinopoulos (1990), Grossman and Helpman (1991) and Aghion and Howitt (1992, henceforth AH).

The AH model is based on the Schumpeterian idea of creative destruction, i.e. the economy is driven by welfare augmenting better products (innovations) and temporary declines (Schumpeter, 1942, Chapter 8). The expected rate of economic growth in AH is determined by the amount of research and its productivity. Innovations are assumed to arrive according to a Poisson process. To quote Aghion and Howitt (1998, p.54):

"When the amount  $n$  is used in research, innovations arrive randomly with a Poisson arrival rate  $\lambda n$ , where  $\lambda > 0$  is a parameter indicating the productivity of the research technology."

This was later also assumed in e.g. Helpman and Trajtenberg (1994) and Maliar and Maliar (2004). The latter study recognizes short waves, but unlike the present study neither accepts negative shocks. There are many real life examples that justify negative shocks, e.g. unsuccessful investments in physical or human capital, bad loans, losses when old investments become worthless and political conflicts. By negative (destructive) random shocks we try to mimic the setbacks in our reduced univariate approach. Moreover, all the models in the quoted works are specified in the time domain, while density distributions are the object of this study.

It is found that some important growth series exhibit heteroscedasticity, which is removed using a filter described in Öller and Stockhammar (2007). The filtered series are shown to be homoscedastic but leptokurticity and positive skewness prevail, rendering a hypothesis of normality dubious.

In the Poisson process the time between each shock is exponentially distributed with intensity  $\lambda$ . In this study however, we have used the exponential distribution to describe the amplitude of the shocks. When  $\lambda$  is small we expect infrequent but large shocks and vice versa. This intuitive way to describe the shocks accords well with modern economic theory. Specifically, to allow for negative or below average shocks, we have used the double exponential (Laplace) distribution obtained as the difference between two exponentially distributed variables with the same value on the parameter  $\lambda$ . The Laplace distribution is symmetric around its mean where the left tail describes below average shocks and vice versa. Due to the asymmetries in these series we have allowed the exponential distribution to take different  $\lambda$ s in the two tails, giving rise to the asymmetric Laplace (AL) distribution. This is our main model candidate. The asymmetric properties of the AL distribution have proved appealing for modelling currency exchange rates, stock prices, interest returns etc. see for instance Kozubowski and Podgorski (1999, 2000) and Linden (2001).

Another plausible explanation is that the long growth series have passed through alternating regimes over the years. Every such regime has its own normal distribution giving rise to a Normal Mixture (NM) distribution, which is our alternative hypothetical distribution, because it is hard to see how this could support the AH hypothesis.<sup>1</sup> The NM distribution, where skewness and leptokurticity are introduced by varying the parameters, was used as early as the late nineteenth century by e.g. Pearson (1895).

It was found that the excess kurtosis in AL models is too large for the filtered (and unfiltered) growth series. The AL (and hence AH) could therefore not be the only source of shocks, so Gaussian noise is introduced. AL distributed innovations are combined with normally distributed shocks leading to the weighted

---

<sup>1</sup>This could also be studied using regime switching models, but given the few observations, we did not pursue this idea.

mixed normal-AL (NAL) distribution. The NAL distribution is capable of generating a wide range of skewness and kurtosis, making the model very flexible. We also consider a convolution of the N and AL distributions. The parameters are estimated using the method of moments (MM).

This paper is organized as follows. The data and the filter are presented in Section 2, and a model discussion together with the proposed model in Section 3. Section 4 contains the estimation set-up and a distributional accuracy comparison. Section 5 concludes.

## 2. The data

In this paper the important US GDP $_q$  (quarterly) 1947-2007, UK GDP $_q$  1955-2007 and the compound GDP 1960-2007 series of the G7 countries<sup>2</sup> are studied as appearing on the websites of Bureau of Economic Analysis ([www.bea.gov](http://www.bea.gov)), National Statistics ([www.statistics.gov.uk](http://www.statistics.gov.uk)) and of OECD ([www.oecd.org](http://www.oecd.org)), respectively. All series are quarterly and seasonally adjusted, final figures, from which we form logarithmic differences, henceforth "growth" for short. In order for the N, NM and NAL distributions to properly describe the rather complex properties, long series are required. The above series are the longest quarterly GDP series available, and the G7 series is based on a large number of observations, albeit not as long as the US and UK ones. The latter series were found significantly heteroscedastic in Öller and Stockhammar (2007). The ARCH-LM test rejected the null hypotheses of homoscedasticity also in the compound G7 GDP series with a  $p$ -value of 0.0004. Heteroscedasticity implies an unequal weighting of the observations leading to inefficient parameter estimates. Heteroscedasticity is removed using the filter in *ibid.*:

$$y_t^* = s_y \left[ \frac{y_t - \sum_{\tau=t-\eta}^{t+\eta} y_\tau / k}{HP^{(\lambda)} \left\{ (2\eta)^{-1} \sum_{\tau=t-\eta}^{t+\eta} \left( y_t - \sum_{\varphi=t-\eta}^{t+\eta} y_\varphi / k \right)^2 \right\}^{1/2}} \right] + \bar{y}, \quad (2.1)$$

where  $y_t = \text{Diff ln GDP}_i$  ( $i = US, UK, G7$ ),  $t = k - \eta, \dots, n - \eta$ ,  $k$  is the window length,  $\eta = (k - 1)/2$ ,  $y_t^*$  is the filtered series and  $s_y$  and  $\bar{y}$  are the estimated standard deviation and arithmetic mean of  $y_t$ , respectively.  $HP^{(\lambda)}$  is the Hodrick-Prescott (1997) filter designed to decompose a macroeconomic time series into a nonstationary trend component and a stationary cyclical residual.

<sup>2</sup> Consists of Canada (1961-), France (1978-), Germany (1991-), Italy (1980-), Japan (1980-), UK (1960-) and US (1960-). Volume national data are scaled up/down to 2000 price levels and the G7 series is calculated as their sum.

Given a seasonally adjusted time series  $z_t$ , the decomposition into unobserved components is

$$z_t = g_t + c_t,$$

where  $g_t$  denotes the unobserved trend component at time  $t$ , and  $c_t$  the unobserved cyclical residual at time  $t$ . Estimates of the trend and cyclical components are obtained as the solution to the following minimization problem

$$\min_{\{g_t\}_{t=1}^N} \left\{ \sum_{t=1}^N c_t^2 + \lambda \sum_{t=3}^N (\Delta^2 g_t)^2 \right\}, \quad (2.2)$$

where  $\Delta g_t = g_t - g_{t-1}$  and  $g_{\min}$  is the HP-filter. The first sum of (2.2) accounts for the accuracy of the estimation, while the second sum represents the smoothness of the trend. The positive smoothing parameter  $\lambda$  controls the weight between the two components. As  $\lambda$  increases, the HP trend become smoother and vice versa. Note that the second sum,  $(\Delta^2 g_t)$ , is an approximation to the second derivative of  $g$  at time  $t$ . For quarterly data (the frequency used in most business-cycle studies) there seems to be a consensus in employing the value  $\lambda = 1600$ . Figure 2.1 shows the Diff ln US, UK and G7 GDP series before and after the filtering<sup>3</sup>.

---

<sup>3</sup>The filtering was done using  $k = 15$  and the standard value for quarterly data,  $\lambda = 1600$ . The bottom panels in Figure 2.1 show the frequency distributions of the unfiltered and filtered data, respectively. The dashed line is the N distribution with the same mean and variance as those of the series. The solid line is the Kernel density estimate

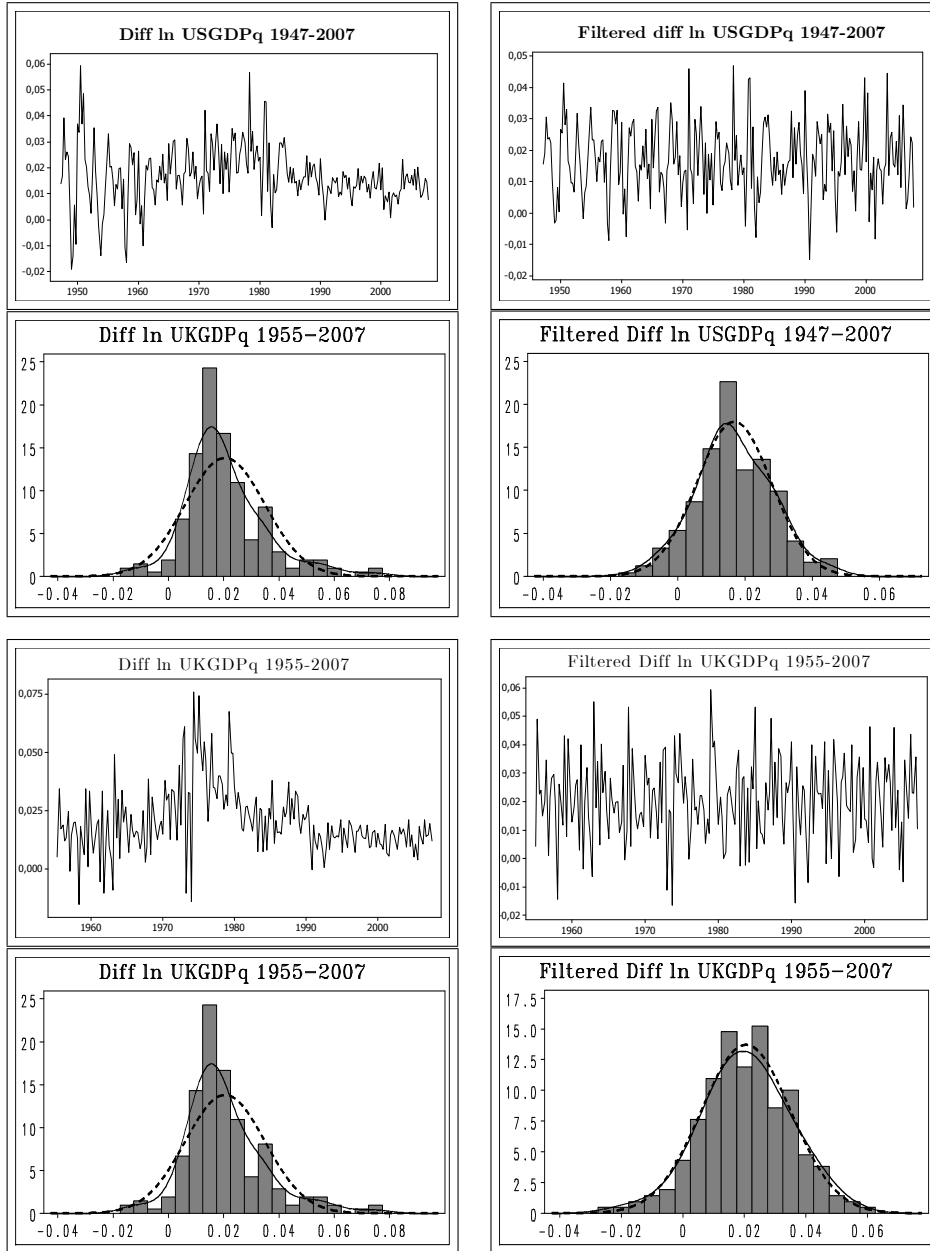
$$\widehat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)$$

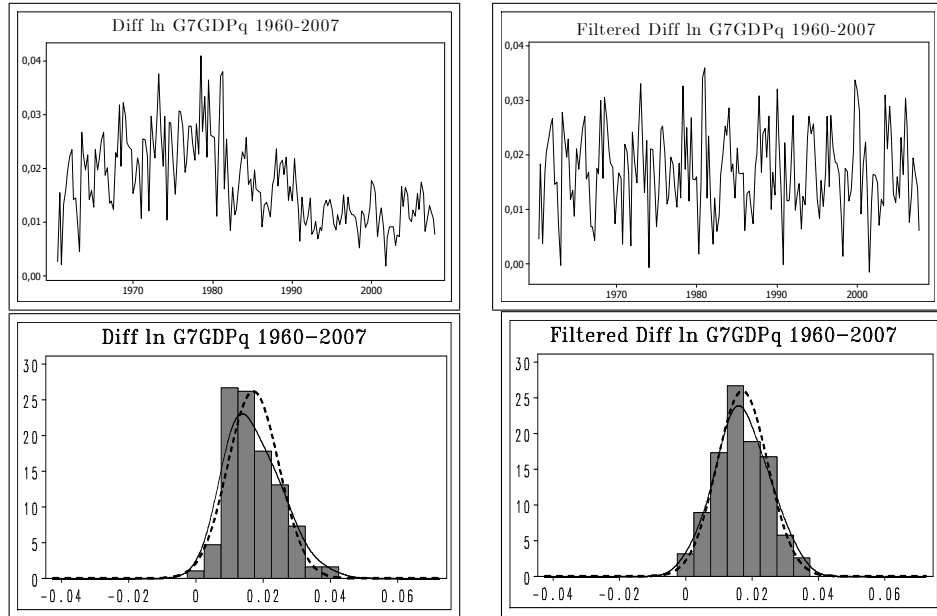
where  $h$  is the bandwidth and  $K(\cdot)$  is the Kernel function. In this study we have used the Gaussian Kernel,  $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$ , and the Silverman (1983) "Rule of Thumb" bandwidth

$$\widehat{h} = \left( \frac{4\widehat{\sigma}^5}{3n} \right)^{1/5} \approx 1.06\widehat{\sigma}n^{-1/5}$$

which is considered to be optimal when data are close to normal.

Figure 2.1: The unfiltered and the filtered Diff ln GDP series





The filter effects on the four moments of US, UK and G7 GDP growth series can be seen in Table 2.1. Period 1 represents the quarters 1947q1-1977q2 (US), 1955q1-1980q2 (UK) and 1960q1-1983q4 (G7). Period 2 contains 1977q3-2007q4 (US), 1980q3-2007q4 (UK) and 1984q1-2007q4 (G7).

Table 2.1: Filter effects on the moments of the Diff ln US, UK and G7 GDP series

	$\hat{\mu}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\sigma}$	$\hat{\tau}$	$\hat{\tau}$	$\hat{\kappa}$	$\hat{\kappa}$
Period	1	2	1	2	1	2	1	2
$y_{t,US}$	0.018	0.016	0.013	0.008	-0.27	1.63	1.00	5.51
$y_{t,US}^*$	0.017	0.017	0.011	0.011	-0.05	0.22	-0.30	0.42
$y_{t,UK}$	0.024	0.016	0.018	0.007	0.45	0.53	0.40	0.49
$y_{t,UK}^*$	0.020	0.020	0.014	0.014	-0.11	-0.16	0.71	-0.35
$y_{t,G7}$	0.022	0.013	0.008	0.004	-0.06	0.51	0.16	0.40
$y_{t,G7}^*$	0.017	0.017	0.008	0.008	-0.01	0.16	-0.31	-0.24

It can be seen that the filter succeeds in stabilizing the means ( $\mu$ ) and the standard deviations ( $\sigma$ ) of the series, and the estimates of skewness ( $\tau$ ) and excess kurtosis ( $\kappa$ ) are also more stable in the filtered series. In Öller and Stockhammar (2007) we showed that this filter does not distort white noise, and thus preserves the dynamics of the time series.

### 3. Model discussion

The unfiltered series in Figure 2.1 do not appear to be normal. Table 3.1 shows that the filter brings them closer to normality

Table 3.1: Filter effects on the normality of the Diff ln US, UK and G7 GDP series

	A-D	S-W	K-S	J-B
$y_{t,US}$	***	***	***	***
$y_{t,US}^*$	*		**	*
$y_{t,UK}$	***	***	***	***
$y_{t,UK}^*$				
$y_{t,G7}$	***	***	***	***
$y_{t,G7}^*$				

In Table 3.1 \*, \*\* and \*\*\* represent significance at the 10%, 5% and 1% levels, respectively, for the null hypothesis of normality. Four commonly used normality tests are reported, where A-D, S-W, K-S and J-B are the Anderson-Darling, Shaphiro-Wilk, Kolmogorov-Smirnov and Jarque-Bera test, respectively. These tests are based on very different measures and can therefore lead to different conclusions.

The failure to reject normality in every entry is surprising if we again take the Figures 2.1 into consideration, where filtered data have fatter tails than the normal distribution. According to e.g. Dyer (1974) the power of these tests is generally low, especially for small samples. Note that the K-S and A-D, J-B statistics for the US reject the null hypotheses of normality. At least for the US series it seems worthwhile to see if there are other distributions that better fit the data. Considering the low power of the tests we will try the same for the UK and the G7 series. The normal distribution remains an alternative hypothesis.

With long time series there is a nonnegligible risk of distributional changes over time. One can argue that data have passed through a number of different regimes, not completely eliminated by the filter (2.1). Every such regime could be N distributed but with different means and variances. The filtered US GDP in Figure 2.1 shows a small hump in the right tail, which may indicate that the data are characterized by at least two regimes, each one N distributed. Given the relatively few observations, the number of regimes is here restricted to two. Moreover, the homoscedasticity test did not detect non-constancy of variances, so even two regimes with different variances could be hard to detect. The introduction of different means and variances for the regimes render it possible to introduce skewness and excess kurtosis in the NM distribution. The probability distribution function (pdf) of the NM distribution is:

$$f_{NM}(y_t^*; \theta) = \frac{w}{\sigma_1 \sqrt{2\pi}} \exp \left\{ -\frac{(y_t^* - \mu_1)^2}{2\sigma_1^2} \right\} + \frac{1-w}{\sigma_2 \sqrt{2\pi}} \exp \left\{ -\frac{(y_t^* - \mu_2)^2}{2\sigma_2^2} \right\}, \quad (3.1)$$



where  $\theta$  consists of the parameters  $(w, \mu_1, \mu_2, \sigma_1, \sigma_2)$  and where  $0 \leq w \leq 1$  is the weight parameter. If this hypothesis provides the best fit, then the AH model gets no support from the data.

Our main hypothesis for growth is a reduced form of the AH model. In this model endogenous growth is driven by creative destruction in which the underlying source is innovations, assumed to be the result of the stochastic arrival of new technologies modelled as a Poisson process. The arrival rate itself is affected by the share of the labour force engaged in research as well as by the Poisson probability of an innovation (research productivity). Each owner of a patent is assumed to have a temporary monopoly of the product lasting until it is replaced and destroyed by a better product.

AH specifies an entire simultaneous model. This paper looks at a reduced, univariate model, where both positive and *negative* shocks hit production exogenously. The drawback with this approach is that the origin of the innovations cannot be identified. The AH model tested in this paper simply assumes that growth is driven by a process, which is the sum of two (one positive, one negative) exponentially distributed random shocks. The distribution of the difference of two exponential random variables is Laplace (L) with pdf:

$$f_L(y_t^*; \theta) = \frac{1}{2\phi} \exp \left\{ -\frac{|y_t^* - \mu|}{\phi} \right\}, \quad (3.2)$$

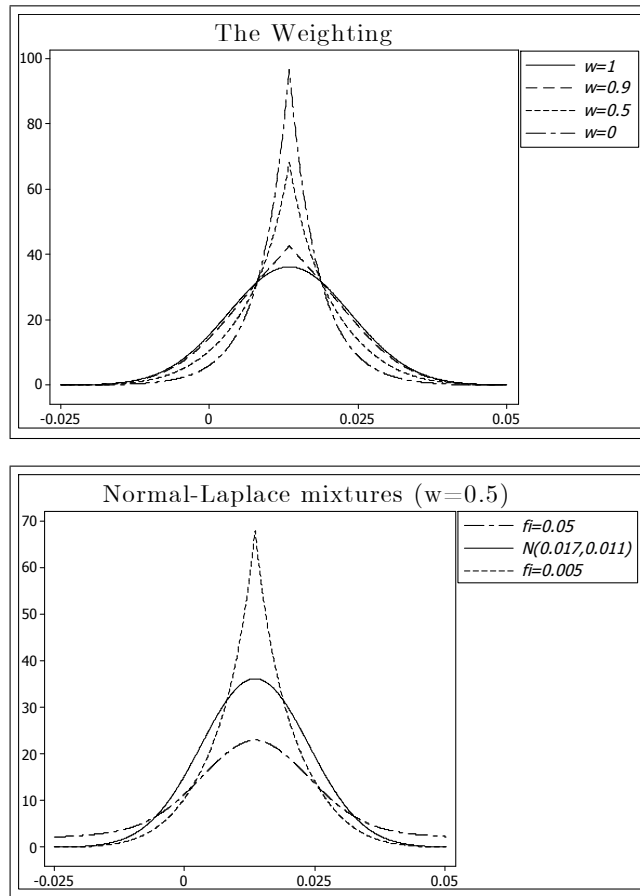
where  $\theta = (\mu, \phi)$ ,  $\mu \in \mathbb{R}$  is the location parameter and  $\phi > 0$  is the scale parameter. The L distribution (which is sometimes also called the double exponential distribution) has been used in many fields: engineering, finance, electronics etc, see Kotz et al. (2001), and the references therein. The L distribution is symmetric around its mean ( $\mu$ ) with  $Var(y) = 2\phi^2$  and excess kurtosis  $\kappa = 3$ . It thus has fatter tails than the N distribution, but it lacks an explicit shape parameter, making it rather inflexible. Also, the excess kurtosis is restricted to the constant value (3), no matter what the kurtosis in the data. Table 2.1 shows the kurtosis in Laplace variables is way too large for the filtered growth series in this study ( $\hat{\kappa} = 0.07$  for the US,  $\hat{\kappa} = 0.19$  for the UK and  $\hat{\kappa} = -0.29$  for the G7). Clearly, the AL as representing the AH model cannot alone explain the data.

The AH model can be modified by allowing it to have a second stochastic component in the sense that its empirical counterpart is buried in Gaussian noise. We thus combine (3.2) with a N distribution via a weight parameter  $w$ . The mixture was introduced by Kanji (1985) to fit wind shear data using the Laplace Normal (LN) mixture distribution specified by:

$$f_{LN}(y_t^*; \theta) = \frac{w}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(y_t^* - \mu)^2}{2\sigma^2} \right\} + \frac{(1-w)}{2\phi} \exp \left\{ -\frac{|y_t^* - \mu|}{\phi} \right\}, \quad (3.3)$$

for  $-\infty < y_t^* < \infty$  and for the parameters:  $-\infty < \mu < \infty$ ,  $0 \leq w \leq 1$  and  $\sigma > 0$ . In (3.3) the L and N distributions have the same mean and variance. This case can be generalized. Jones and McLachlan (1990) assumed different variances and showed that this leads to an even better fit than Kanji's. The characteristics of this density are shown in Figure 3.1.

Figure 3.1: LN densities



The upper panel in Figure 3.1. shows different weightings of the two components in the Normal-Laplace mixture distribution (with  $\mu=0.017$ ,  $\sigma=0.011$ ,  $\phi=0.005$ ). The lower panel shows the pure  $N(0.017, 0.011)$  distribution together with two mixture distributions with  $w=0.5, \phi=0.05$  and  $\phi=0.005$  respectively.

The above L and LN mixture distributions do not account for skewness in the data. A suitable skewed generalization of the L distribution is presented in

McGill (1962) who proposes an Asymmetric Laplace (AL) distribution of the form

$$f_{AL}(y_t^*; \boldsymbol{\theta}) = \begin{cases} \frac{1}{2\psi} \exp\left\{\frac{y_t^* - \mu}{\psi}\right\} & \text{if } y_t^* \leq \mu \\ \frac{1}{2\phi} \exp\left\{\frac{\mu - y_t^*}{\phi}\right\} & \text{if } y_t^* > \mu \end{cases}, \quad (3.4)$$

where again  $\mu$  is the location parameter, for which the median is the Maximum Likelihood (ML) estimate, and  $\boldsymbol{\theta} = (\mu, \phi, \psi)$ . This distribution is negatively skewed if  $\psi > \phi$ , and vice versa for  $\psi < \phi$ . If  $\psi = \phi$  the AL collapses to the L distribution. In AL,  $\psi$  is the parameter of shocks weaker than the trend and  $\phi$  that of stronger shocks than the trend. If  $\psi \neq \phi$  then Schumpeterian shocks that lead to weaker than trend growth behave differently from stronger growth shocks. During the last couple of decades, various forms and applications of AL distributions have appeared in the literature, see Kotz et al. (2001) for an exposé. Linden (2001) used an AL distribution to model the returns of 20 stocks, where  $\psi$  and  $\phi$  were shown to be highly significant. Another recent paper is Yu and Zhang (2005) who used a three-parameter AL distribution to fit flood data.

The AL arises as the limiting distribution of the random sum (or difference) of independent and identically, exponentially distributed random variables with finite variances. An advantage of the AL distribution is that, unlike the L distribution, the kurtosis is not fixed. The AL distribution is even more leptokurtic than the L distribution with an excess kurtosis that varies between three and six (the smallest value for the L distribution, and the largest value for the exponential distribution). The main advantage of the AL distribution is that it is skewed (for  $\psi \neq \phi$ ), conforming with the empirical evidence in Figure 2.1. To this distribution we are adding Gaussian noise. To the authors' best knowledge this distribution has not been used before for macroeconomic time series data. We assume that each shock is an independent drawing from either a N or an AL distribution. The probability density distribution of the filtered growth series ( $y_t^*$ ) can then be described by a weighted sum of N and AL random shocks, i.e:

$$f_{NAL}(y_t^*; \boldsymbol{\theta}) = \frac{w}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y_t^* - \mu)^2}{2\sigma^2}\right\} + (1-w) \begin{cases} \frac{1}{2\psi} \exp\left\{\frac{y_t^* - \mu}{\psi}\right\} & \text{if } y_t^* \leq \mu \\ \frac{1}{2\phi} \exp\left\{\frac{\mu - y_t^*}{\phi}\right\} & \text{if } y_t^* > \mu \end{cases}, \quad (3.5)$$

where  $\boldsymbol{\theta}$  consists of the five parameters ( $w, \mu, \sigma, \phi, \psi$ ). Model (3.5) is referred to as the mixed normal-asymmetric Laplace (NAL) model. Note that (as in Jones and McLachlan, 1990) equal means but unequal variances are assumed for the components in the proposed distribution. It has a jump discontinuity at  $\mu$  when  $\psi \neq \phi$ , see Figure 3.2. Looking at the smoothed empirical distributions in Figure 2.1, the discontinuity seems counterintuitive. However, the histograms in Figure 2.1 lend some support to a jump close to  $\mu$ . The cumulative distribution function (cdf) of 3.5 is given in the Appendix. Figure 3.2 shows NAL densities

for three different values of the weight parameter  $w$ .

Figure 3.2: NAL densities

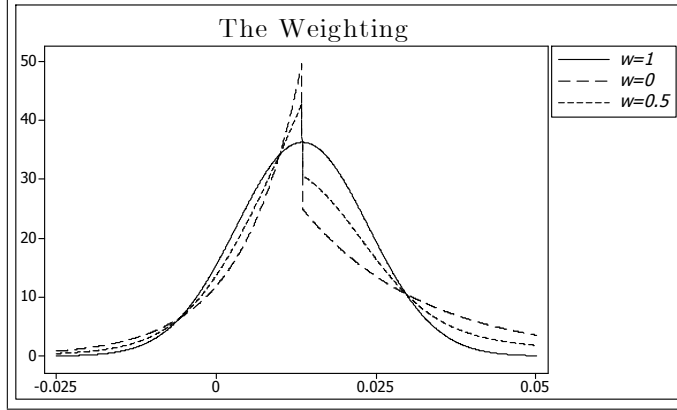


Figure 3.2 shows a pure  $N(0.017,0.011)$ , an  $AL(0.01,0.02,0.017)$  distribution ( $w=1$  and  $w=0$  respectively) and a compound of these two components with  $w=0.5$ . Note the discontinuity at  $\mu$ .

The AH hypothesis also accords with the assumption that each shock is a random mixture of a N and an AL distributed component. We then arrive at the convoluted version suggested by Reed and Jorgensen (2004). Instead of using the AL parameterization in (3.4) they used:

$$f_{AL^*}(y_t^*; \boldsymbol{\theta}) = \begin{cases} \frac{\alpha\beta}{\alpha+\beta} \exp\{\beta y_t^*\} & \text{if } y_t^* \leq 0 \\ \frac{\alpha\beta}{\alpha+\beta} \exp\{-\alpha y_t^*\} & \text{if } y_t^* > 0 \end{cases} \quad (3.6)$$

which was convoluted with a N distribution giving the following pdf:

$$f_{c-NAL}(y_t^*; \boldsymbol{\theta}) = \frac{\alpha\beta}{\alpha+\beta} \phi\left(\frac{y_t^* - \mu}{\sigma}\right) \left[ R\left(\alpha\sigma - \frac{y_t^* - \mu}{\sigma}\right) + R\left(\beta\sigma + \frac{y_t^* - \mu}{\sigma}\right) \right] \quad (3.7)$$

where  $\boldsymbol{\theta} = \alpha, \beta, \mu, \sigma$  and  $R(z) = \frac{\Phi^c(z)}{\phi(z)}$  is the Mill's ratio. Ibid. called this distribution the four parameter Normal Laplace distribution, but it is here called the convoluted NAL, c-NAL. As can be seen from Figure 3.3 this distribution lacks the jump in the mode.

Figure 3.3: *c-NAL densities*

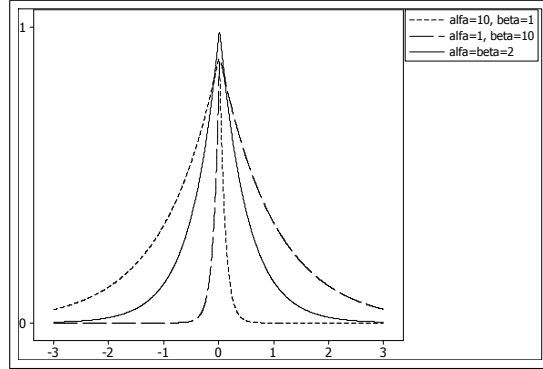


Figure 3.3 shows *c-NAL* densities using  $\mu=0.017, \sigma=0.011$ ,  $(\alpha, \beta)=(10, 1), (1, 10)$  and  $(\alpha, \beta)=(2, 2)$ , respectively.

The *c-NAL* distribution has the advantage of being more parsimonious than *NAL*. Whether it is more suitable to describe economic growth is the issue of the next Section.

## 4. Estimation and distributional accuracy

In this section we will fit all four distributions (*N*, *NM*, *NAL* and *c-NAL*) in order to find out which one best describes the data. The five parameters in the *NM* distribution (3.1) will be estimated using the method of moments (MM) for the first four moments of the same. A close distributional fit is important in density forecasting. As noted by e.g. Fryer and Robertson (1972), the method of maximum likelihood breaks down for this model. The noncentral moments of (3.1) are given in the Appendix. Equating the theoretical and the observed first four moments using the five parameters yields infinitely many solutions<sup>4</sup>. A way around this dilemma is to fix  $\mu_1$  to be equal to the observed mode, which is here approximated by the maximum value of the Kernel function of the empirical distribution ( $\max f_K(y_i^*)$ ). In the presence of large positive skewness we can expect  $\mu_1$  to be smaller than  $\mu$ , and vice versa. Here  $\hat{\mu}_{1,US}$ ,  $\hat{\mu}_{1,UK}$  and  $\hat{\mu}_{1,G7}$  are substituted for  $\max f_K(y_{US}^*) = 0.0142$ ,  $\max f_K(y_{UK}^*) = 0.0196$  and  $\max f_K(y_{G7}^*) = 0.0159$ . The observed moments and the corresponding MM parameter estimates for the filtered series (using the above values for  $\mu_{1,US}$ ,  $\mu_{1,UK}$  and  $\mu_{1,G7}$ ) are given in Table 4.1.

<sup>4</sup>We tried to make use of the fifth moment, but in none of the series did it even at a 10% significance level differ from zero in a  $\chi^2(1)$  test.

Table 4.1: Sample moments and estimated parameters of the NM assumption<sup>5</sup>

	Sample noncentral moments				Estimated NM parameters			
	$E(y_t^*)$	$E(y_t^{*2})$	$E(y_t^{*3})$	$E(y_t^{*4})$	$\hat{w}$	$\hat{\mu}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$
US	0.0168	0.0004	$1.1 \cdot 10^{-5}$	$3.4 \cdot 10^{-7}$	0.833	0.030	0.010	0.005
UK	0.0205	0.0006	$2.1 \cdot 10^{-5}$	$8.1 \cdot 10^{-7}$	0.919	0.030	0.015	0.011
G7	0.0171	0.0003	$8.0 \cdot 10^{-6}$	$2.0 \cdot 10^{-7}$	0.866	0.024	0.008	0.006

The log likelihood function of the NAL distribution is:

$$\ln L(\boldsymbol{\theta}) = w \left[ -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{\sum_{t=1}^n (y_t^* - \mu)^2}{2\sigma^2} \right] \\ + (1-w) \left[ -\left( n \ln(2\psi) + \frac{1}{2\psi} \sum_{t=1}^n (y_t^* - \mu) \right) \times I(y_t^* \leq E(y_t^*)) \right] \\ - \left( n \ln(2\phi) + \frac{1}{2\phi} \sum_{t=1}^n (\mu - y_t^*) \right) \times I(y_t^* > E(y_t^*)) \right]$$

However, ML is in this case cumbersome, if at all possible, so again the parameters are estimated by the method of moments. This also enables a fair comparison with the NM and the c-NAL distributions. The formulae of the noncentral and central moments of (3.5) are given in the Appendix. There are five parameters and only four moment conditions, so equating the theoretical and the observed first four moments will not give a unique solution. We now fix  $\mu$  to be equal to the ML estimate with respect to  $\mu$  in the AL distribution, that is the observed median,  $\widehat{md}$ . Here  $\hat{\mu}_{US}$ ,  $\hat{\mu}_{UK}$  and  $\hat{\mu}_{G7}$  are substituted for  $\widehat{md}_{US} = 0.0156$ ,  $\widehat{md}_{UK} = 0.0203$  and  $\widehat{md}_{G7} = 0.0167$ . The parameter values that satisfy the moment conditions are:

Table 4.2: Estimated NAL parameters

	Estimated parameters			
	$\hat{w}$	$\hat{\sigma}$	$\hat{\psi}$	$\hat{\phi}$
US	0.711	0.012	0.006	0.014
UK	0.828	0.015	0.018	0.017
G7	0.939	0.008	0.008	0.020

Table 4.2 shows that the Gaussian noise component dominates. In the US and G7 series  $\hat{\psi}$  is much smaller than  $\hat{\phi}$ , which indicates that growth shocks that are weaker than trend have a smaller spread than above trend shocks. Together with a mean growth larger than zero this ensures long-term economic growth.

Reed and Jorgensen (2004) provided some guidelines on how to estimate the c-NAL parameters in (3.7) using ML techniques. To be consistent and to make fair comparisons, the parameters are here again estimated using MM. Ibid. also supplied the first four cumulants, and in order to find the MM parameter

<sup>5</sup>The MM technique does not provide standard errors for the estimates.

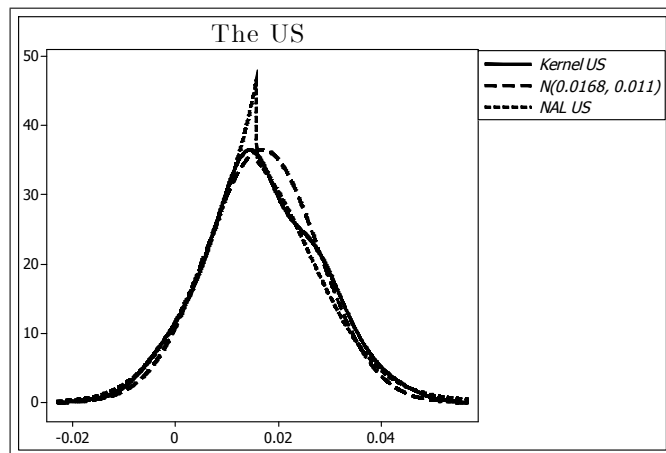
estimates we provide the first four noncentral moments in the Appendix. Note that the c-NAL distribution has four parameters so there is no need to fix one of the parameters to find a unique MM solution. The MM estimates are given in Table 4.3, which shows that c-NAL does not capture the asymmetry ( $\hat{\alpha}$  is very close to  $\hat{\beta}$ ), cf. Table 4.2.

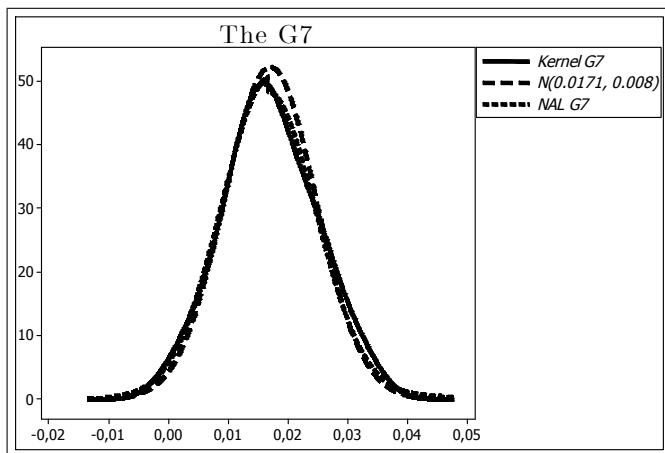
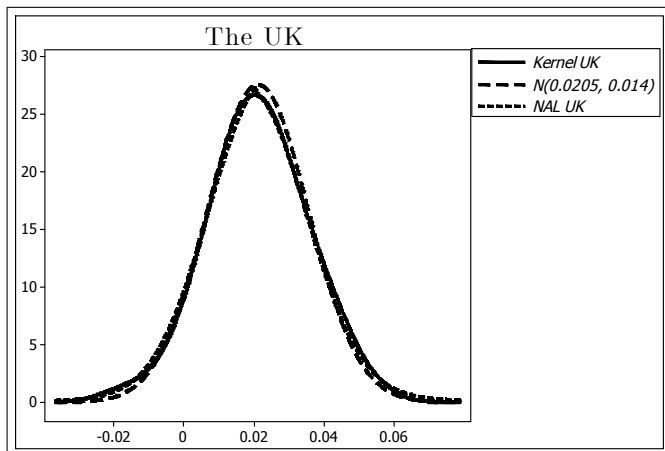
Table 4.3: Estimated c-NAL parameters

	Estimated parameters			
	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\beta}$
US	0.017	0.009	197.0	198.0
UK	0.021	0.010	140.1	136.0
G7	0.017	0.007	498.2	498.8

Figure 4.1 shows the estimated NAL and c-NAL distributions together with the benchmark N distributions and the smoothed empirical GDP series. The graphs, especially of the US distributions, reveal the discontinuity at  $\hat{\mu}$ , resulting in a poignant peak, as discussed in connection with Figure 3.2.

Figure 4.1: Distributional comparison





We have chosen to compare the ordinates of the empirical distributions and the hypothetical N, NM, NAL and c-NAL distribution for 1000 points, however dropping points outside the interval  $(\hat{\mu} - 4\hat{\sigma}, \hat{\mu} + 4\hat{\sigma})$ . Table 4.4 shows the distributional fit for the three hypothesis. Four accuracy measures are used. The Root Mean Square Error, RMSE, is here defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{1000} [f_K(y_i^*) - \hat{f}(y_i^*)]^2}{1000}},$$

where  $f_K(y_i^*)$  is the Kernel function of the empirical distribution, and  $\hat{f}(y_i^*)$  is the hypothetical distribution. The sum is taken over ordinates of equidistant points on the horizontal axis, and hence there are more points where the distributions are almost parallel to the x-axis, providing these points with more weight. Thus, the peak to the left of the median will considerably affect RMSE



for the US.

Percentage error measures are widely used but they also have their disadvantages. They are undefined at  $f_K(y_i^*) = 0$ , and they have a very skewed distribution for  $f_K(y_i^*)$  close to zero. The median absolute percentage error measure, MdAPE, defined as:

$$MdAPE = median \left( \frac{100 \times |f_K(y_i^*) - \hat{f}(y_i^*)|}{f_K(y_i^*)} \right),$$

is here, because of the asymmetry, a better measure than its close relative, the MAPE (defined as  $\frac{100}{n} \times \sum_{i=1}^n |f_K(y_i^*) - \hat{f}(y_i^*)| / f_K(y_i^*)$ ). Yet another disadvantage is that positive errors are counted heavier than negative ones. This is the reason why so-called "symmetric" measures have been suggested (Makridakis, 1993). One is the Symmetric Median Absolute Percentage Error, sMdAPE

$$sMdAPE = median \left( \frac{200 \times |f_K(y_i^*) - \hat{f}(y_i^*)|}{f_K(y_i^*) + \hat{f}(y_i^*)} \right).$$

Hyndman and Koehler (2006) suggested the Mean Absolute Scaled Error, MASE, defined as:

$$MASE = \frac{1}{1000} \left| \frac{f_K(y_i^*) - \hat{f}(y_i^*)}{\frac{1}{999} \sum_{i=2}^{1000} |f_K(y_i^*) - f_K(y_{i-1}^*)|} \right|.$$

Ibid. showed that this measure is widely applicable and less sensitive to outliers and small samples than the other measures.

All the above four measures are reported in Table 4.4.

Table 4.4: Distributional accuracy comparison

		RMSE	MdAPE	sMdAPE	MASE
US	$N(0.0168, 0.0111)$	1.76	16.65	16.73	15.38
US	NM	<b>1.41</b>	13.98	13.19	15.59
US	NAL	1.73	<b>9.07</b>	<b>9.12</b>	<b>12.68</b>
US	c-NAL	1.59	14.25	14.43	14.11
UK	$N(0.0205, 0.0145)$	0.65	13.00	13.63	9.46
UK	NM	0.59	15.46	14.99	8.66
UK	NAL	<b>0.45</b>	<b>9.47</b>	<b>9.50</b>	<b>6.77</b>
UK	c-NAL	0.58	15.02	14.80	7.99
G7	$N(0.0171, 0.0077)$	2.21	18.69	20.13	14.60
G7	NM	1.64	11.07	10.55	11.78
G7	NAL	<b>1.20</b>	<b>8.65</b>	<b>8.99</b>	<b>9.00</b>
G7	c-NAL	1.70	12.27	11.61	12.55

The NAL distribution using the parameter values in Table 4.2 is superior to the N, NM and the c-NAL distribution according to every measure, except RMSE for the US where, as expected, the discontinuity peak has a large impact on the measure. NAL shows on average 27.6%, 29.2% and 48.3% better fit for the US, UK and G7, respectively (comparing with the benchmark N distribution). Comparing to the NM distribution, NAL is an improvement with on average 15.5% , 30.2% and 21.7% for the US, UK and G7. Finally, the NAL is on average a 18.8%, 27.6% and a 27.4% improvement over the c-NAL distribution. According to this numerical comparison, the US, UK and G7 GDP series could be looked upon as samples from a NAL distribution with the parameter estimates in Table 4.2. In other words, the AH hypothesis of economic growth could be correct, if we accept that shocks are either Poisson (AL) *or* N distributed, with N dominating.

Kernel estimation is based on subjective choices both of function and of bandwidth. But so are goodness of fit tests and it is well known that tests based on both approaches have low power. To be on safer ground we have chosen also to test the histograms using three different numbers of bins. We then get unique critical values for these tests which enable calculation of p-values. Table 4.5 reports the p-values of the  $\chi^2$  tests using 10, 15 and 20 bins when testing the null hypotheses  $H_{0,1} : y^* \sim N$ ,  $H_{0,2} : y^* \sim NM$ ,  $H_{0,3} : y^* \sim NAL$  and  $H_{0,4} : y^* \sim c-NAL$ .

Table 4.5: The  $\chi^2$  goodness of fit test

Bins	$H_{0,1}: y^* \sim N$			$H_{0,2}: y^* \sim NM$			$H_{0,3}: y^* \sim NAL$			$H_{0,4}: y^* \sim c-NAL$		
	10	15	20	10	15	20	10	15	20	10	15	20
US	0.03	0.01	0.04	0.13	0.05	0.15	0.20	0.22	0.27	0.06	0.04	0.07
UK	0.91	0.70	0.83	0.84	0.34	0.57	0.71	0.64	0.32	0.90	0.72	0.88
G7	0.18	0.28	0.22	0.17	0.25	0.05	0.37	0.47	0.22	0.21	0.33	0.19

While the power of these tests is low, they still indicate that the NAL distribution fits growth best.

## 5. Conclusions

The hypothesis that economic growth could be described by a reduced Poisson model is not strongly contradicted by data. The Laplace distribution is unable to describe the asymmetric and just slightly leptokurtic shape. A mixed Normal-Asymmetric Laplace (NAL) distribution is introduced and is shown to better describe the density distribution of growth than the Normal, Normal Mixture, convoluted NAL and Laplace distributions. This paper thus, from a new angle, supports the hypothesis that innovations arriving according to a Poisson process play an important role in economic growth, as suggested by e.g. Helpman and Trajtenberg (1994) and Maliar and Maliar (2004). But one has to accept that most shocks are Gaussian. Since according to the AH hypothesis,  $\lambda$  measures the productivity of only positive shocks (research productivity), the  $\phi$  parameter could be given the same interpretation. On the other hand,  $\psi$  would measure the strength of regimes of low productivity or downright destructive shocks. Thus, our technique provides a way to estimate these qualities, and perhaps to compare different economies.

NAL implies a decomposition of the shocks into an AL and an N part. The mean, variance, skewness and the fatness of the tails stand in relation to the five parameters in the NAL distribution, and are estimated using MM on the first four moments. The moment generating function and the first four central and noncentral moments of the NAL distribution are provided. Because of the close distributional fit, the NAL distribution is a good choice for density forecasting of GDP growth series, or for that matter of any series with these features. The NAL distribution could also be used for conditional density forecasts applying priors on the parameters  $\phi$  and  $\psi$ , but that merits another study.

## Acknowledgments

This research was supported by the Department of Statistics at Stockholm University, Royal Swedish Academy of Sciences, the International Institute of Forecasters and by the Societas Scientiarum Fennica. We gratefully acknowledge helpful comments from Daniel Thorburn of Stockholm University and from Roy Batchelor and Lorenzo Trapani of Cass Business School, London. Parts of this paper have been presented at the International Symposium on Forecasting in 2007 and 2008 and also at various other venues. We are grateful for the many suggestions from seminar participants. Finally, we would like to thank an associate editor and two referees for very constructive criticism and comments, which have led to substantial improvements of the paper.

## Appendix. Theoretical moments

The noncentral moments of the Normal mixture distribution (3.1) are given by

$$E(Y^n) = \frac{w}{\sigma_1\sqrt{2\pi}} \sum_{k=0}^n \binom{n}{k} \{1 + (-1)^k\} \mu_1^{n-k} 2^{(k-1)/2} \sigma_1^{k+1} \Gamma\left(\frac{k+1}{2}\right) + \frac{1-w}{\sigma_2\sqrt{2\pi}} \sum_{k=0}^n \binom{n}{k} \{1 + (-1)^k\} \mu_2^{n-k} 2^{(k-1)/2} \sigma_2^{k+1} \Gamma\left(\frac{k+1}{2}\right)$$

and specifically the first four moments are

$$\begin{aligned} E(Y) &= w\mu_1 + (1-w)\mu_2 \\ E(Y^2) &= w(\mu_1^2 + \sigma_1^2) + (1-w)[\mu_2^2 + \sigma_2^2] \\ E(Y^3) &= w(\mu_1^3 + 3\mu_1\sigma_1^2) + (1-w)[\mu_2^3 + 3\mu_2\sigma_2^2] \\ E(Y^4) &= w(\mu_1^4 + 6\mu_1^2\sigma_1^2 + 3\sigma_1^4) + (1-w)[\mu_2^4 + 6\mu_2^2\sigma_2^2 + 3\sigma_2^4]. \end{aligned}$$

The cdf of the NAL distribution (3.5) is given by:

$$F(y_t^*; \theta) = \begin{cases} w\Phi\left(\frac{y_t^* - \mu}{\sigma}\right) + \frac{1-w}{2} \exp\left\{\frac{y_t^* - \mu}{\psi}\right\} & \text{if } y_t^* \leq \mu \\ w\Phi\left(\frac{y_t^* - \mu}{\sigma}\right) + (1-w)\left(1 - \frac{1}{2} \exp\left\{\frac{y_t^* - \mu}{\phi}\right\}\right) & \text{if } y_t^* > \mu \end{cases},$$

where  $\Phi(\cdot)$  denotes the cdf of the standard normal distribution.

The noncentral moments of (3.5) are given by

$$E(Y^n) = \frac{w}{\sigma\sqrt{2\pi}} \sum_{k=0}^n \binom{n}{k} \{1 + (-1)^k\} \mu^{n-k} 2^{(k-1)/2} \sigma^{k+1} \Gamma\left(\frac{k+1}{2}\right) + (1-w) \left[ \frac{1}{2\psi} \sum_{k=0}^n \binom{n}{k} (-1)^k \mu^{n-k} \psi^{k+1} k! + \frac{1}{2\phi} \sum_{k=0}^n \binom{n}{k} \mu^{n-k} \phi^{k+1} k! \right].$$

Specifically,

$$\begin{aligned} E(Y) &= w\mu + (1-w) \left[ \mu + \frac{\phi - \psi}{2} \right] \\ E(Y^2) &= w(\mu^2 + \sigma^2) + (1-w) [\mu^2 + \psi(\psi - \mu) + \phi(\phi + \mu)] \\ E(Y^3) &= w(\mu^3 + 3\mu\sigma^2) + (1-w) \left[ \mu^3 + \frac{3\psi}{2} (2\mu\psi - 2\psi^2 - \mu^2) + \frac{3\phi}{2} (\mu^2 + 2\mu\phi + 2\phi^2) \right] \\ E(Y^4) &= w(\mu^4 + 6\mu^2\sigma^2 + 3\sigma^4) + (1-w) [\mu^4 + 2\psi(3\mu^2\psi - 6\mu\psi^2 + 6\psi^3 - \mu^3) + 2\phi(\mu^3 + 3\mu^2\phi + 6\mu\phi^2 + 6\phi^3)]. \end{aligned}$$

Note that the mean of the NAL distribution can be written as

$$E(Y) = \mu + (1 - w) \left[ \frac{\phi - \psi}{2} \right],$$

which clearly shows the obvious fact that positively skewed distributions ( $\phi > \psi$ ) will have a mean larger than  $\mu$  and vice versa. The central moments are

$$\begin{aligned} E(Y) &= 0 \\ Var(Y) &= w\sigma^2 + (1 - w) \left[ \frac{\phi\psi(1 - w)}{2} + \frac{(\phi^2 + \psi^2)(3 + w)}{4} \right] \\ \tau(Y) &= \frac{1}{4} [7\phi^3 + 3\phi^2\psi - 3\phi\psi^2 - 7\psi^3] - \frac{3w(\phi - \psi)}{4} [(\phi + \psi)^2 + 2\sigma^2] \\ &\quad - \frac{w^2(\phi - \psi)}{4} [(\phi + \psi)^2 - 2\sigma^2] - \frac{w^3(\phi - \psi)^3}{4} \\ \kappa(Y) &= \frac{3}{16} [39\phi^4 + 20\phi^3\psi + 10\phi^2\psi^2 + 20\phi\psi^3 + 39\psi^4] \\ &\quad - \frac{3w}{4} [5\phi^4 + 8\phi^3\psi + 5\psi^4 - 2\phi^2\psi^2 - 4\sigma^4 + \phi^2(6\psi^2 - 2\sigma^2) \\ &\quad + 4\phi\psi(2\phi^2 + \sigma^2)] - \frac{3w^2(\phi - \psi)^2}{8} [7\phi^2 + 10\phi\psi + 7\psi^2 + 8\sigma^2] \\ &\quad - \frac{3w^3(\phi - \psi)^2}{4} [(\phi + \psi)^2 - 2\sigma^2] - \frac{3w^4(\phi - \psi)^4}{16}. \end{aligned}$$

The first four noncentral moments of the c-NAL (3.7) are:

$$\begin{aligned} E(Y) &= \mu + \frac{1}{\alpha} + \frac{1}{\beta} \\ E(Y^2) &= \mu^2 + \sigma^2 - \frac{2\mu}{\beta} + \frac{2(\mu\beta - 1)}{\alpha\beta} + \frac{2}{\alpha^2} + \frac{2}{\beta^2} \\ E(Y^3) &= \frac{1}{\alpha^3\beta^3} [6\beta^3 + 6\alpha\beta^2(\beta\mu - 1) + 3\alpha^2\beta(2 - 2\beta\mu + \beta^2(\mu^2 + \sigma^2)) \\ &\quad + \alpha^3(6\beta\mu - 6 - 3\beta^2(\mu^2 + \sigma^2) + \beta^3(\mu^3 + 3\mu\sigma^2))] \\ E(Y^4) &= \frac{1}{\alpha^4\beta^4} [24\beta^4 + 24\alpha\beta^3(\beta\mu - 1) + 12\alpha^2\beta^2(2 - 2\beta\mu + \beta^2(\mu^2 + \sigma^2)) \\ &\quad + 4\alpha^3\beta(6\beta\mu - 6 - 3\beta^2(\mu^2 + \sigma^2) + \beta^3(\mu^3 + 3\mu\sigma^2)) \\ &\quad + \alpha^4(24 - 24\beta\mu + 12\beta^2(\mu^2 + \sigma^2) - 4\beta^3(\mu^3 + 3\mu\sigma^2)) \\ &\quad + \beta^4(\mu^4 + 3\sigma^4 + 6\mu^2\sigma^2)] \end{aligned}$$

## References

- Aghion, P. and Howitt, P. (1992) A model of growth through creative destruction. *Econometrica*, 60, 323-351.
- Aghion, P. and Howitt, P. (1998) *Endogenous growth theory*. The MIT press, Cambridge.
- Dyer, A. R. (1974). Comparisons of tests for normality with a cautionary note. *Biometrika*, 61, 185-189.
- Fryer, J. G. and Robertson, C. A. (1972) A comparison of some methods for estimating mixed normal distributions, *Biometrika*, 59, 639-648.
- Grossman, G. and Helpman, E. (1991) Quality ladders in the theory of growth. *Review of Economic Studies*, 63, 43-61.
- Helpman, E. and Trajtenberg, M. (1994) A Time to Sow and a Time to Reap: Growth Based on the General Purpose Technologies. Centre for Economic Research Policy, working paper no. 1080.
- Hodrick, R. J. and Prescott, E. C. (1997) Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29, 1-16.
- Hyndman, R. J. and Koehler, A. B. (2006) Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, 679-688.
- Jones, P. N. and McLachlan, G. J. (1990) Laplace-normal mixtures fitted to wind shear data. *Journal of Applied Statistics*, 17, 271-276.
- Kanji, G. K. (1985) A mixture model for wind shear data. *Journal of Applied Statistics*, 12, 49-58.
- Kotz, S. Kozubowski, T. J. and Podgorski, K. (2001) *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Birkhäuser, Boston.
- Kozubowski, T. J. and Podgorski, K. (1999) A class of asymmetric distributions. *Actuarial Research Cleraring House*, 1, 113-134.
- Kozubowski, T. J. and Podgorski, K. (2000) Asymmetric Laplace distributions. *The Mathematical Scientist*, 25, 37-46.
- Linden, M. (2001) A model for stock return distribution. *International Journal of Finance and Economics*, 6, 159-169.
- Lucas, R. E. (1988) On the Mechanics of Economic Development. *Journal of Monetary Economy*, 22, 3-42.
- Makridakis, S. (1993) Accuracy measures: theoretical and practical concerns. *International Journal of Forecasting*, 9, 527-529.
- Maliar, L. and Maliar, S. (2004) Endogenous growth and endogenous business cycles. *Macroeconomic Dynamics*, 8, 559-581.

- McGill, W. J. (1962) Random fluctuations of response rate. *Psychometrika*, 27, 3-17.
- Öller, L-E. and Stockhammar, P. (2007) A heteroscedasticity removing filter. *Research Report 2007:1*, Department of Statistics, Stockholm university.
- Pearson, K. (1895) Contributions to the mathematical theory of evolution, 2: skew variation. *Philosophical Transactions of the Royal Society of London (A)*, 186, 343-414.
- Reed, W. J. and Jorgensen, M. A. (2004) The double Pareto-lognormal distribution - A new parametric model for size distributions. *Communications in Statistics: Theory and Methods*, 33(8), 1733-1753.
- Romer, P. M. (1986) Increasing Returns and Long-Run Growth. *Journal of Political Economy*, 94, 1002-1037.
- Schumpeter, J. A. (1942) *Capitalism, Socialism and Democracy*. Harper, New York.
- Segerstrom, P. Anant, T. Dinopoulos, E. (1990) A Shumpeterian model of the product life cycle. *American Economic Review*, 80, 1077-1091.
- Silverman, B. W. (1986) *Density estimation for statistics and data analysis*. Chapman and Hall, London.
- Yu, K. and Zhang, J. (2005) A three-parameter asymmetric Laplace distribution and its extension. *Communications in Statistics-Theory and Methods*, 34, 1867-1879.