University of Leicester

Department of Economics

Choosing the Right Skew Normal Distribution: the Macroeconomist' Dilemma

Wojciech Charemza, University of Leicester

Carlos Díaz, University of Leicester

Svetlana Makarova, University College London

Working Paper No. 15/08

May 2015

CHOOSING THE RIGHT SKEW NORMAL DISTRIBUTION: THE MACROECONOMIST' DILLEMA

WOJCIECH CHAREMZA*, CARLOS DÍAZ **
AND SVETLANA MAKAROVA***

May 2015

This paper has been developed from its earlier version 'Too many skew normal distributions: the practitioner's dilemma'; University of Leicester Working Paper No 13/07.

* University of Leicester, UK
** University of Leicester, UK
*** University College London, UK

KEYWORDS: skew normal distributions, ex-post uncertainty, inflation forecasting, economic policy

JEL codes: E17, C46, E52, E37

ACKNOWLEDGEMENT

Financial support of the ESRC/ORA project RES-360-25-0003 *Probabilistic Approach to Assessing Macroeconomic Uncertainties* is gratefully acknowledged. This research used the ALICE High Performance Computing Facility at the University of Leicester. We are very grateful to the participants of the *Third International Symposium in Computational Economics and Finance* (ISCEF), Paris, April 2014 for their comments and suggestions. We are solely responsible for all remaining deficiencies.

ABSTRACT

The paper discusses the consequences of possible misspecification in fitting skew normal distributions to empirical data. It is shown, through numerical experiments, that it is easy to choose a distribution which is different from that which generated the sample, if the minimum distance criterion is used. The distributions compared are the two-piece normal, weighted skew normal and the generalized Balakrishnan skew normal distribution which covers a variety of other skew normal distributions, including the Azzalini distribution. The estimation method applied is the simulated minimum distance estimation with the Hellinger distance. It is suggested that, in case of similarity in values of distance measures obtained for different distributions, the choice should be made on the grounds of parameters' interpretation rather than the goodness of fit. For monetary policy analysis, this suggests application of the weighted skew normal distribution, which parameters are directly interpretable as signals and outcomes of monetary decisions. This is supported by empirical evidence of fitting different skew normal distributions to the *ex-post* monthly inflation forecast errors for Poland, Russia, Ukraine and U.S.A., where estimations do not allow for clear distinction between the fitted distributions for Poland and U.S.A.

1. INTRODUCTION

During the last decade a substantial development of the theory and applications of skew normal distributions, which contain normal distribution as their special symmetric case, can be observed. The first distribution of this kind applied in empirical macroeconomics was probably the so-called two-piece normal (or split normal) distribution, TPN, originated by John (1982) and developed further by Kimber (1985). It gained substantial popularity among the practitioners; in particular it has been widely used by economic forecasters for constructing probabilistic forecasts of inflation (see e.g. seminal paper by Wallis, 1999). Further breakthrough was made by Azzalini (1985, 1986), who developed a theory of univariate, and then multivariate, skew normal distributions. These distributions have been recently subject of substantial generalisations. Most notably, the Balakrishnan skew normal distribution has been proposed by Sharafi and Behboodian (2008), generalized Balakrishnan skew normal distribution, GBSN, by Yagedari, Gerami and Khaledi (2007), and developed further by Hasanalipour and Sharafi (2012), Fujisawa and Abe (2012), Mameli and Musio (2013), and others.

Such plethora of distributions to choose from provides a practitioner with a dilemma of which one to choose. Most intuitively, the distribution to be selected is the one with the best fit to the data. But what if the distributions considered fits to the data equally good (or equally bad), so that they are statistically undistinguishable? Leaving aside the numerical problems which might affect our decision (for their discussion see e.g. Pewsey, 2000, Monti, 2003, and Castro, San Martín and Arellano-Valle 2008 and Franceschini and Loperfido, 2014) we argue that in selecting appropriate skew normal distribution the secondary criterion of choice, in case there the selection cannot be clearly decided on the statistical grounds, it should be done on the basis of interpretation of the parameters of the skew normal distributions. In some cases, the customised distributions can be derived, with parameters directly related to the particular theory or the phenomenon described. In particular, Charemza, Díaz and Makarova (2014) proposed a skew normal distribution, called weighted skew normal distribution, WSN, which, if applied for modelling macroeconomic uncertainty, has parameters directly interpretable in the context monetary policy.

In order to tackle the problems formulated above, we have decided to put the WSN, TPN and GBSN distributions to a goodness of fit contest. These distributions are described in a greater detail in Section 2. Section 3 explains general settings and estimation procedure. Section 4 presents the results of a Monte Carlo study evaluating the probabilities of choosing a wrongly specified skew normal distribution on the basis of its fit. It concludes that it might be indeed difficult to distinguish between the WSN, TPN and GBSN distributions on the grounds of their fit. Section 5 shows empirical results of estimation of skew normal distributions for the four-step ahead inflation forecast uncertainty for Poland, Russia, Ukraine and U.S.A.. Indeed, for Poland and U.S.A., where all distributions fit well, it is not practically possible to decide which distribution is the right one. This, however, can be decided by choosing the distribution with economically interpretable parameters which, in this case, is WSN. Section 6 concludes.

2. THREE SKEW NORMAL DISTRIBUTIONS

There are three distributions which we consider in this paper: weighted skew normal, WSN (which we regard as the benchmark one), two-piece normal, TPN, and the Yagedari, Gerami and Khaledi (2007) generalized Balakrishnan skew normal distribution, GBSN.

A special case of the random variable Z with WSN distribution, as defined by Charemza, Díaz and Makarova, (2014), is:

$$Z = X + \alpha \cdot Y \cdot I_{Y>\tau_{up}} + \beta \cdot Y \cdot I_{Y<\tau_{low}} \quad , \tag{1}$$

where:

$$I_{Y>\tau_{up}} = \begin{cases} 1 & \text{if } Y > \tau_{up} \\ 0 & \text{otherwise} \end{cases}, \; I_{Y<\tau_{low}} = \begin{cases} 1 & \text{if } Y < \tau_{low} \\ 0 & \text{otherwise} \end{cases}, \; (X,Y) \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}\right),$$

$\tau_{low} < \tau_{up}$; $\alpha, \beta \in \mathbb{R}$; $\sigma^2 \in \mathbb{R}^+$; and $|\rho| \le 1$.

Following Charemza, Díaz and Makarova (2014), if WSN is fitted to data expressing macroeconomic uncertainty (e.g. *ex-post* baseline forecast errors of inflation or output). It is assumed that the baseline forecast can be in turn improved by second stage forecast, given by *Y*. Under these settings the parameters of WSN have the following interpretation:

(i) α and β represent the marginal effect of the stimulative and contractionary economic policies on uncertainty respectively;

(ii) τ_{low} and τ_{up} represent the thresholds deciding about the relevance of the second stage forecast information for the policy decisions;

(iii) σ^2 that is variance of *X* and *Y*, represent the uncertainty related to the second stage forecast information used for improving the baseline forecast outcome;

(iv) ρ, that is the correlation coefficient between *X* and *Y*, describes the accuracy of the second stage forecasts.

A random variable with TPN distribution is defined by its *pdf*:

$$f_{TPN}(t;\sigma_1,\sigma_2,\mu) = \begin{cases} A\exp\left\{-(t-\mu)^2 / 2\sigma_1^2\right\} & \text{if } t \le \mu \\ A\exp\left\{-(t-\mu)^2 / 2\sigma_2^2\right\} & \text{if } t > \mu \end{cases}, \quad t \in \mathbb{R} \quad ,$$

where $A = \sqrt{2/\pi} \cdot (\sigma_1 + \sigma_2)^{-1}$. Three parameters to be estimated are $\sigma_1, \sigma_2 \in \mathbb{R}^+$ and $\mu \in \mathbb{R}$.

It is often interpreted in the context of forecast uncertainty as a representation of the balance of risks the over-and underestimated forecasts (see Wallis, 2004).

The third distribution considered here, the GBSN, is given by the following *pdf*:

$$f_{GBSN}(t;n,m,\delta) = \frac{1}{C(n,m,\delta)}[\Phi(\delta t)]^n[1-\Phi(\delta t)]^m \varphi(t) \quad , \quad t \in \mathbb{R} \quad ,$$

where $C(n,m,\delta) = \sum_{i=0}^{m} \binom{m}{i} (-1)^i \int_{-\infty}^{\infty} [\Phi(\delta t)]^{n+i} \varphi(t) dt$, Φ and φ are respectively the *cdf* and *pdf* of the standard normal distribution and the non-negative integers *n* and *m* and $\delta \in \mathbb{R}$ are the parameters. The GSBN includes the Balakrishnan skew normal distribution for *m* = 0, and the original Azzalini skew normal distribution (with the probability density function $f_{\mathrm{SN}}(t;\lambda) = 2\varphi(t)\Phi(\lambda t)$) for *n* = 1 and *m* = 0). Azzalini distribution is also a special case of the WSN for, $\lambda = -2\rho$, $\tau_{up} = \beta = 0$ and $\sigma^2 = 1$. All three distributions can be reduced to a standard normal: WSN for $\alpha = \beta = 0$ and $\sigma^2 = 1$; TPN for $\sigma_1 = \sigma_2 = 1$ and $\mu = 0$; GBSN for $n = 1$ and $\delta = m = 0$ or $n = m = 0$. So far, the parameters of the GBSN distribution have not been given any particular interpretation.

Figure 1 compares the *pdf*'s of WSN, TPN and GBSN with approximately identical first three moments, namely with means equal to 0.78, variances equal to 0.39 and the coefficient of skewness equal to 0.85. For comparison, the normal distribution with mean equal to 0.78 and variance of 0.86 is plotted in the background.

Figure 1: *pdf*'s of WSN, TPN and GBSN with identical first three moments mean= 0.78, variance=0.39, coef. of skewness=0.85.

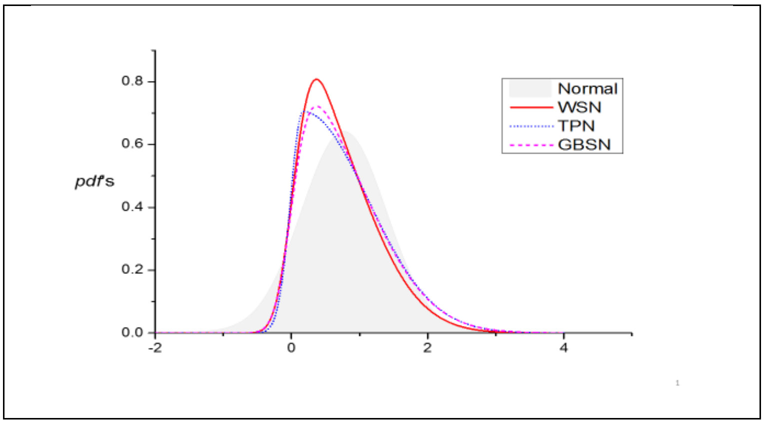
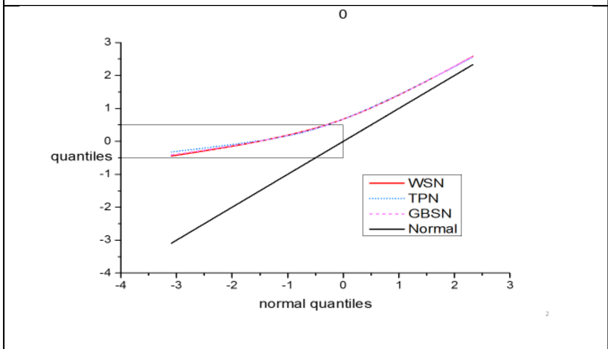
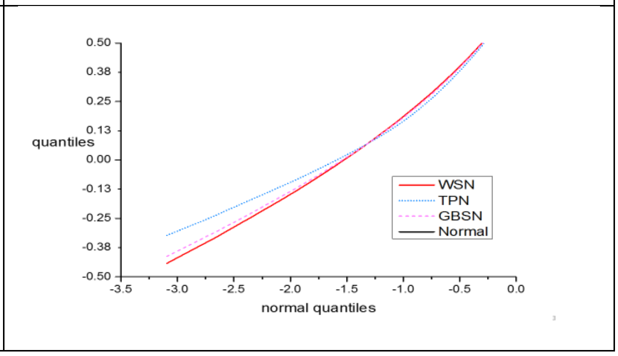


Figure 2 shows the Q-Q diagram, that is a scatter diagram of the quantiles of each of the distributions depicted at Figure 1 against the normal distribution. Figure 2a gives the Q-Q plot for the entire range of quantiles, from 0.01 to 0.99, and Figure 2a gives a close-up of Figure 2a for the left tails of the distributions

Figure 2: Q-Q plots of WSN, TPN and GBSN distributions

Fig 2a: full range of quantiles	Fig. 2b: a closeup of the left tails
	

Figures 1-2 illustrate potential problems in distinguishing between the distributions. The GBSN and TPN have, in the case illustrated, nearly identical modes, with TPN having a slightly thicker right tail. The corresponding quintiles are nearly identical, with the exception of the left-hand side quantiles, where the GBSN are slightly lower than the other corresponding quantiles, relatively to the identical quantile of the normal distribution. Nevertheless, these differences are small, which suggest practical problems in discrimination between skew normal distributions.

3. ESTIMATION AND GENERAL SETTINGS

Estimation of WSN, TPN and GBSN distributions by the maximum likelihood or the generalized method of moments is numerically awkward. This problem is particularly well discussed for the Azzalini distribution (see e.g. Azzalini and Capitanio, 1999, Sartori, 2006, Franceschini and Loperfido, 2014), and is evident also for all three families of distributions considered here. For this reason we have resorted to simulation-based estimation methods. These methods are particularly attractive as it is straightforward to derive random number generators for all three distributions. For WSN given by (1) it is described in Charemza, Díaz and Makarova (2014), for TPN in Nakatsuma (2003) and for GBSN in Yagedari, Gerami and Khaledi (2007). With the use of these generators and inspired by Greco (2011) we have applied the simulated minimum distance estimators method (*SMDE,* see Charemza *et al*., 2012), which consists of fitting the approximated by simulation density function to empirical histograms of data and applying a minimum distance criterion.

The version of *SMDE* applied here can be defined as:

$$\hat{\omega}_n^{SMDE} = \arg\min_{\omega\in\Omega} \left\{ \mu_w \left(d(g_n\ ,\ f_{t,\omega}) \right)_{t=1}^{T} \right\} \quad ,$$

where $f_{t,\omega}$ is the approximation of the *pdf* , f_ω, of a random variable obtained by generating t = 1,…,T replications (drawings) from a distribution with parameters ω ($\omega \in \Omega \subset \mathbb{R}^k$), g_n denotes the density of empirical sample of size *n*, μ_w is an operator based on *T* replications, which deals with the problem of the ‘noisy’ criterion function (median, in this case), and $d(\bullet,\bullet)$ is the distance measure. The minimum distance measures, *MD*, applied here are that of the Cressie and Read (1984) power divergence disparities family given by:

$$d(g_n\ ,\ f_{t,\omega}) = \frac{1}{\lambda_{CR}(\lambda_{CR}+1)} \sum_{i=1}^{m+1} g_n(i) \left[\left(\frac{g_n(i)}{f_{t,\omega}(i)} \right)^{\lambda_{CR}} - 1 \right] \quad , \tag{2}$$

where *m* denotes the number of cells in which data are organized. For $\lambda_{CR} = 1$ formula (2) gives the Pearson χ^2 measure, for $\lambda_{CR} = -1/2$ the (twice squared) Hellinger distance (*HD*) and for $\lambda_{CR} = -2$ the Neyman χ^2 measure. For $\lambda_{CR} \to 0$ and $\lambda_{CR} \to -1$ the continuous limits of the right-hand side expression in (2) are respectively the likelihood disparity (*LD*) and the Kullback-Leibler divergence statistics. Cressie and Read (1984) advocate optimal setting $\lambda_{CR} = 3/2$.[1] More details and the properties of the MSD are discussed in Charemza et al (2012).

4. FIT OF TRUE AND FALSE MODELS

As the initial objective of this paper is to check whether using the best fit criterion for selecting the best type of a skew normal distribution might lead to choosing a false one, we have set up three data generating processes (*DGP*’s, or ‘true models’) and fitted all three distributions to the generated data.

The *DGP*’s are:

DGP 1: WSN with $\alpha = -2.0$, $\beta = -0.5$, $\sigma^2 = 1$, $\tau_{up} = -\tau_{low} = 1$ and $\rho = 0.75$.

DGP 2: TPN with $\sigma_1 = 1.5$, $\sigma_2 = 0.5$, $\mu = 0.4$.

DGP 3: GBSN with n = 2, k = 1 and $\delta = -0.3$.

All three *DGP*’s have similar first three moments, as given in Table 1:[2]

---

[1] For a complex discussion and alternatives see Basu, Shioya and Park (2011).

[2] Computing moments of GBSN requires numerical integration over an infinite interval. The algorithms applied here are that of Sikorski and Stenger (1984), named inthp1and inthp2 in GAUSS 13 and later versions.

Table 1: Mean, st. deviation and skewness of *DGP*'s

	Mean	st. dev.	Skewness
DGP 1	-0.363	1.069	-0.628
DGP 2	-0.398	1.113	-0.695
DGP 3	-0.207	0.925	-0.687

For each *DGP*, and for sample sizes of 100, 150, 200, 250, 300, 350, 400, 450 and 500, there have been generated *Nrepl* = 1,000 replications. For each simulated sample we have fitted all three distributions using the *SMDE* method outlined in Section 3. Problem in comparison arises due to the fact that WSN is a 5-parameters distribution and TPN and GBSN have 3 parameters each. In order to allow for a fair comparison, we have decided to keep the ρ parameter constant (that is, $\rho = 0.75$). Also, we are keeping the threshold parameters, τ_{up} and τ_{low} constant in estimation, albeit in two different variations. In the first variation, denoted by WSN(0), we keep the thresholds fixed as in the *DGP* 1, that is $\tau_{up} = -\tau_{low} = 1$. In the second variation, we made the thresholds dependent on σ in such way that $\tau_{up} = \sigma$ and $\tau_{low} = -\tau_{up}$. We denote this as WSN(1). Hence we are left with three parameters to estimate: α, β, and σ. In such settings the skewness in WSN is induced only through the differences between the parameters α and β. Such settings are close to that applied in the empirical models analysed in Section 5.

As a simple, naïve, misspecification measure, we use the frequency of cases when $d_0(\xi^i) > d_1(\xi^i)$, where d_0 denotes the minimum distance measure computed for the estimated properly specified distribution in the i^{th} replication ξ^i, and d_1 denotes the minimum distance measure computed for one of the misspecified distribution estimated using the same generated data. That is, we do the comparison in pairs, comparing the properly specified distribution with the falsely specified one. By the properly specified distribution we understand the distribution of the same type as used for generating the sample. The distance criterion used here is the twice squared Hellinger distance, *HD* (results for other criteria are available on request; they do not differ much from these presented in this paper).

Another misspecification measure is based on bootstrapping the ratios distance measures for two alternative distributions fitted to the same sample. We have used methodologies developed originally for comparing variances: simple bootstrap and Efron bootstrap (see e.g. Sun, Chernick and LaBudde, 2011).

The algorithm for the simple bootstrap is the following:

Step 1: Draw *M* pairs of $\{d_0(\xi^k), d_0(\xi^j)\}$, *k*,*j* = 1,…, 1,000, $k \neq j$. *M* should be large, e.g. 10,000;

Step 2: Compute the ratio of distance measures $r_0^h = \dfrac{d_0(\xi^k)}{d_0(\xi^j)}$, *h* = 1,2,…,*M*;

Step 3: Compute the 95$^{\text{th}}$ quantile of the distribution of r_0^h denoted as $q_{0.95}$;

Step 4: Check the simulated bootstrap criterion for the case where $d_0(\xi^i) > d_1(\xi^i)$ as:

$$\frac{d_1(\xi^i)}{d_0(\xi^i)} > q_{0.95} \quad .$$

The frequency of cases where the above inequality is fulfilled tells about the probability of undertaking the right decisions regarding the distribution by rejecting the wrong one. It approximates the probability of rejecting the null hypothesis that the distance measures for the true and false distributions are identical with the implicit alternative that the distribution on which $d_1(\xi^i)$ is based is false. Consequently, the higher is this ratio, the false distribution is chosen less often. Efron bootstrap is similar, except that in Step 1 drawing of pairs is made from a set of all $d_0(\xi^k)$, $d_1(\xi^k)$ rather than from $d_0(\xi^k)$ alone. Results in this case are more robust, as the equality of the distance measures is explicit under the null.

Tables 2, 3 and 4 present respectively the naïve misspecification measure and also those based on the simple and Efron bootstraps and for twice squared Hellinger minimum distance criterion. Results for other criteria and different sample sizes are available on request.

Table 2 Frequency of cases where $d_0(\xi^i) > d_1(\xi^i)$

Sample size	*DGP* 1 (WSN)		*DGP* 2: (TPN)			*DGP* 3: (GBSN)		
	TPN	GBSN	WSN (1)	WSN (0)	GBSN	WSN (1)	WSN (0)	TPN
100	0.380	0.479	0.504	0.507	0.396	0.314	0.287	0.269
250	0.261	0.229	0.37	0.442	0.091	0.304	0.299	0.228
500	0.258	0.06	0.186	0.341	0.005	0.377	0.353	0.233

Table 3 Simple simulated bootstrap power

Sample size	*DGP* 1 (WSN)		*DGP* 2: (TPN)			*DGP* 3: (GBSN)		
	TPN	GBSN	WSN(1)	WSN(0)	GBSN	WSN(1)	WSN(0)	TPN
100	0.088	0.045	0.047	0.046	0.026	0.186	0.207	0.219
250	0.135	0.099	0.073	0.046	0.049	0.201	0.191	0.267
500	0.148	0.216	0.137	0.071	0.183	0.152	0.168	0.254

Table 4 Efron simulated bootstrap power

Sample size	*DGP* 1 (WSN)		*DGP* 2: (TPN)			*DGP* 3: (GBSN)		
	TPN	GBSN	WSN(1)	WSN(0)	GBSN	WSN(1)	WSN(0)	TPN
100	0.085	0.069	0.045	0.038	0.074	0.071	0.093	0.086
250	0.117	0.112	0.086	0.061	0.156	0.09	0.081	0.116
500	0.131	0.161	0.124	0.103	0.174	0.076	0.082	0.113

Tables 2-4 show that results of fitting WSN and TPN to data generated from GBSN behave differently to that fitted to data generated from WSN or TPN distributions. Let us first concentrate on evaluating the misspecification in case when data are generated by WSN and TPN; it is clearly difficult to distinguish between these two distributions. Using *MD* criterion for the small sample size it is practically haphazard to find out which statistic is smaller regardless of the data generating process. In particular, if data are generated from TPN, there is a virtually equal chance that WSN would fit better than the true TPN distribution. However, with the increase in sample size the frequencies of cases where the *MD* statistics for the 'true' distribution is smaller than for the 'false' one increase, suggesting the consistency of choice based on the *MD* criterion. This is confirmed by the bootstrap results. The empirical power of the tests based on the *MD* statistics is, in absolute terms, not high. Even for samples of size 500 it is not reaching 20%. In another words, it is in practice problematic to distinguish between the WSN and TPN distributions.

Nevertheless, some differences between the fits given by WSN and TPN can be observed here. Generally TPN is more often falsely well approximated by WSN, particularly in the case when $\tau_{up} = \sigma$ that is for WSN, than WSN by TPN. Also, for middle-sized samples (250 observations) chances for proper identification of WSN against TPN by rejecting the null of identical *MD* statistics are visibly higher than otherwise, albeit still small in absolute terms.

For data generated by WSN and TPN, the danger of misspecification by falsely fitting GBSN is visibly smaller. Except for small samples of data generated by WSN, *MD* statistics for GBSN are usually bigger than for two remaining distributions in this case than the corresponding WSN and TPN statistics, reducing the chance of distributional misspecification. Also, the empirical power of the *MD* ratio test rises relatively quickly with an increase in sample size exceeding, in some cases, 20% for large samples.

In contrast to WSN and TPN, data generated by GBSN exhibit different patterns. In terms of power of the bootstrap tests, they can also be easily confused with two other distributions as the power of the *MD* ratio test is low. However, the power of the test is not visibly increasing with the increase of sample size, causing doubts regarding the consistency. On the positive side, the naïve misspecification benchmark based on the differences between the *MD* statistics for the true and false distributions is less often false than in the case of data generated from WSN and TPN.

To sum up, one would expect the confusions in deciding which skew normal distribution is the best one o be relatively frequent, if the selection is based on the goodness of fit measures alone.

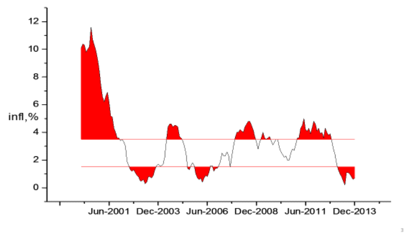
5. EMPIRICAL RESULTS: DISTRIBUTIONS OF INFLATION FORECAST ERRORS FOR POLAND, RUSSIA, UKRAINE AND U.S.A.

The distributions discussed above have been used for the evaluation of the probabilities of headline (CPI) inflation being within target bands for Poland, Russia, Ukraine and U.S.A. in the middle of 2013, with the use of data ending four months earlier, which is in February 2013. Economies of these countries differ substantially between themselves, both in term of economic growth, average level of inflation and types of conducted monetary policy. Among these countries only Poland conducted a reasonably successful monetary policy since 1998, with a clearly defined inflation target set at 2.5%, with ±1% band, since 2004. Russia, although officially pursuing inflation stabilization, was, in fact, targeting exchange rate stabilisation, which resulted in inflation fuelled by a 'dirty float' (see Vdovichenko and Voronina, 2004). Since 2010, it has been implementing inflation targeting more efficiently. For 2013, the year of the forecast, official inflation target was 5%-6%. Nevertheless, in order to allow for a comparison with other countries, we have evaluated the probability of inflation being within the ±1% band around its upper limit, that of around 6%. Ukraine monetary policy was the least transparent. It first had implemented the exchange rate targeting which was followed, since 2000, by the exchange rate pegging. In 2013 it announced transition to inflation targeting, with indications that a likely target will be 5%. Consequently, we have assumed for Ukraine the target band of ±1% around 5%. For the analysis of the development of monetary policy in these three countries see e.g. Égert and MacDonald (2008). In the U.S.A., the nominal target of 2% has been announced by the Federal Open Market Committee in early 2012. Before that, the unofficial target was within the range of 1.7%-2%. In line with the assumptions made for other countries, for U.S.A. we are assuming the target of 2%, with the band of ±1%. Figure 3 shows the inflation data for all four countries with the inflation target bands depicted. It depicts the differences in inflation dynamics in the analysed countries and also the differences in the frequencies the inflation target bands for 2013 were crossed in the past.

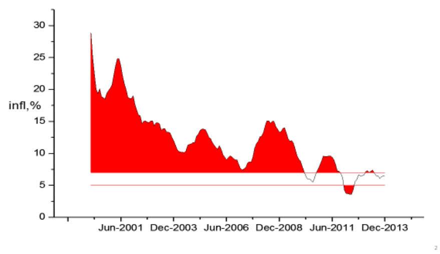
We have computed the probabilities of inflation being within the bands using the distributions of 4-step ahead forecast errors obtained for forecasts made prior to February 2013, approximated by the skew normal distributions discussed above, that is WSN, TPN and GBSN. The practice of using past forecast errors for approximating forecast uncertainties has often been used before, especially by the central banks' practitioners (see e.g. Kowalczyk, 2013). Even if not used directly, there is a practice of comparing the distributions of point forecast errors with the distributions used in probabilistic forecasting (Hall and Mitchell, 2007, Dowd, 2008). The alternative would be to apply the distributions derived from the surveys of professional forecasters (see Clements, 2014; Lahiri and Sheng, 2010; Lahiri, Peng and Sheng, 2014). However, systematic data for forecasts' surveys are available only for Poland and U.S.A., and their reliability for producing probabilistic is yet to be proven (see Bowles et al., 2007, Andrade and Bihan, 2013).

Figure 3: Annual inflation in Poland, Russia, Ukraine and U.S.A., 2000-2013, and the assumed target bands for 2013

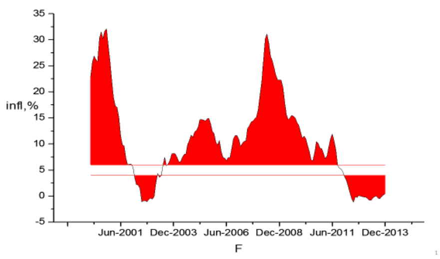
Poland



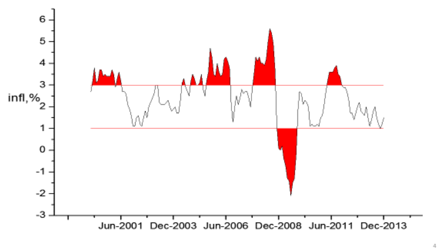
Russia



Ukraine



U.S.A.



After checking for the order of seasonal and non-seasonal integration by the Taylor (2003) test which takes into account the possibility of the presence of unit roots at frequencies other than tested, we have estimated the seasonal ARIMA (SARIMA) model for inflation data y_t:

$$\phi(L)\Psi(L^s)\Delta^{\kappa}\Delta^{D}y_t = \theta(L)\Theta(L^s)u_t \ ,$$

where L is the lag operator, κ is the order of integration of the regular part of y_t, $\Delta^{\kappa} = (1-L)^{\kappa}$ is the regular difference operator of order κ, $\Delta^{D} = (1-L^s)^{D}$ is the seasonal difference operator of order s for a seasonal $I(D)$ process and u_t is the error term. Polynomials ϕ, θ, Ψ and Θ are based on regular (L) and seasonal (L^s) lag operators correspondingly. Their orders have been obtained using the Gómez and Maravall (1998) procedure which is based on an automatic lag selection criterion which leads to a minimum of Ljung-Box the autocorrelation statistic. The entire data span, for the annual inflation recorded monthly in percentages, is from August 1994 (September 1994 for Russia) to February 2012.[3] Due to data availability, for Ukraine we have used a slightly shorter data span, from January 1995 to February 2012. Out of sample forecasts has been computed recursively, starting from the

[3] Data are from the official statistical agencies of each country, available at http://www.tradingeconomics.com

initial period and updating the sample by one observation in each recursion. Initial (for first recursion) period for estimation has been defined as a maximum of the first 80 observations of the series. Basic descriptive statistics of the recursive forecast errors are given in Table 5.

Table 5: Basic characteristics of 4-step ahead forecast errors

	Poland	Russia	Ukraine	U.S.A.
Full data span	07-1994 02-2013	08-1994 02-2013	01-1995 02-2013	07-1994 02-2013
total no. of observation in sample.	222	221	217	222
no. of obs. for estimation of the densities	138	137	133	138
mean.	-0129	-0.187	0.106	0.034
std. dev.	1.037	1.977	3.560	1.379
skewness	0.211	0.793	0.106	0.533

All three skew normal distributions discussed here, which is WSN, TPN and GBSN, have been fitted to the forecast errors data. As in Section 4, we have estimated three parameters of WSN: α, β, and σ, keeping $\rho = 0.75$, $\tau_{up} = \sigma$ and $\tau_{low} = -\tau_{up}$.

Table 6 presents the results of estimation of the parameters of the density functions. As in the previous section, for each distribution three parameters have been estimated by the *SMDE*. As parameters *m* and *n* of the GBSN distribution are integers, their standard errors have not been computed. For the non-integer parameters standard errors are given in brackets below the estimates.

Table 6: Results of empirical estimation of different skew normal distributions

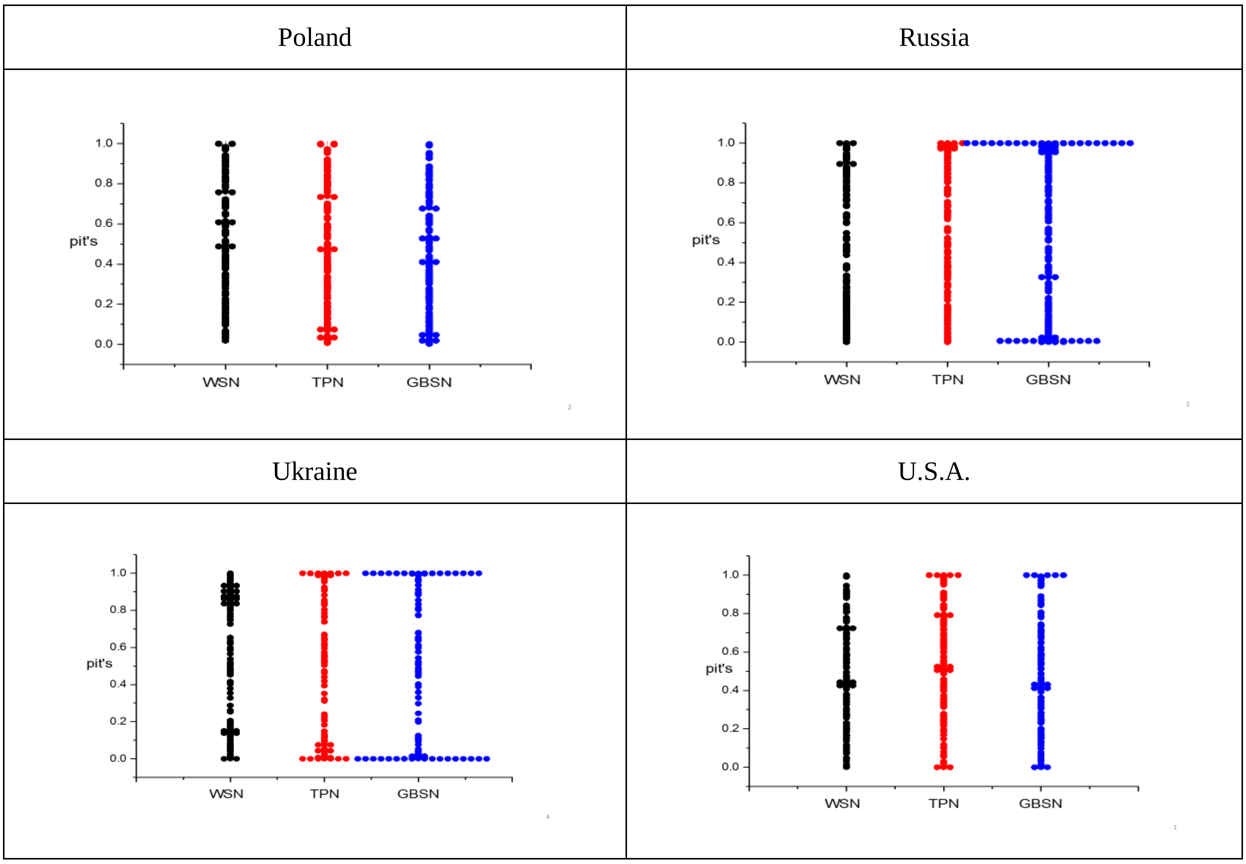
	Parameters	Poland	Russia	Ukraine	U.S.A.
WSN	α	-1.817 (0.1803)	-2.281 (0.1303)	-3.320 (0.1262)	-1.887 (0.1051)
	β	-1.223 (0.1792)	-2.358 (0.6400)	-3.478 (0.372)	-2.241 (0.0042)
	σ	1.000 (0.001)	0.999 (0.0.002)	0.999 (0.004)	0.851 (0.0290)
	MD	2.629	21.05	47.42	1.19
TPN	σ_1	1.169 (0.1101)	1.400 (0.617)	1.995 (0.0144)	0.902 (0.0533)
	σ_2	0.8422 (0.367)	2.000 (0.001)	2.000 (0.001)	1.306 (0.0712)
	μ	0.088 (0.482)	-0.7222 (0.4990)	-0.153 (0.275)	-0.336 (0.303)
	MD	3.10	1.11	51.52	6.26
GBSN	*n*	5	5	5	5
	m	2	1	5	2
	δ	-0.061 (0.0510)	-0.084 (0.0110)	0.000 (0.09800)	-0.976009 (0.0229)
	MD	4.71	72.80	251.7	8.90

The distance measure criterion suggests the choice of different distributions for particular countries. For Poland the best fit is that of TPN, followed closely by WSN, and for U.S.A the best fit is that by WSN. As concluded in Section 4, there is high chance of distributional misspecification between WSN and TPN. With this is mind and taking into account that, for

Polish data, differences in *MD*'s in fitting WSN and TPN are not negligible, we can interpret parameters of WSN in the light of monetary policy outcomes. For Poland, the positive difference between the absolute values of the estimates of α and β in WSN indicates footprints of the prevalence of anti-inflationary policy over the output-stimulating policy. For U.S.A., however, where such difference is negative, there is some evidence of signs of output-stimulating policy. For Russia, TPN gives the best fit, and for Ukraine all minimum distance statistics are rather large, suggesting poor fit of all the distributions considered.

In order to discriminate between the distributions further, we have tested which of these distributions fits better to the observed data with the use of the probability integral transform (*pit*) test (see Diebold, Gunther and Tai, 1998; for application to evaluation of inflation probabilistic forecast see Clemens, 2004, and Galbraith and van Norden, 2012); for other similar approaches and applications to inflation modelling see Mitchell and Hall (2005). The probability integral transform is defined as the probability of observing values of a random variable not greater than its realized value. If the forecasted density is close enough to the true but unknown density, *pit's* will be uniform on the interval from zero to one. If several *pit*'s (that is, for different forecasts) are available, one can test their accuracy by checking whether their values are uniformly distributed using well known 'goodness-of-fit' tests. Figure 4 give the scatter diagram of *pit*'s for all three distributions and countries analysed and Table 7 gives the results of the Cramer-von Mises test for uniformity of *pit*'s.[4]

Figure 3: *pit*'s for fitted skew normal distributions



[4] Another test often used for evaluating the uniformity of pit's is that of Berkovitz (2001). Hovever, we have decided not to use it, as this test is proved to be biased in evaluation of multi-step forecasts (Dowd, 2007).

Table 7: Test statistic: Cramer-von Mises statistics for testing uniformity
significant statistics are marked by * for 10% significance and ** for 5% significance

	WSN	TPN	GBSN
Poland	0.034	0.100	0.611**
Russia	0.671**	0.328	2.492**
Ukraine	0.739**	1.649**	2.633**
U.S.A.	0.095	0.279	0.068

Both Figure 3 and Table 7 confirms better fit of all distributions for the Poland and U.S.A. than for two remaining countries. In particular, the uniformity of GBSN estimates is questionable, with *pit*'s concentrated close to zero and one.

Finally, Table 8 gives the probabilities that inflation, in July 2013, that is four month's periods after the end of the sample data, will be within the target intervals. They were obtained by computing numerical integrals of the respective estimated *pdf*'s over the target interval. It also presents *pit*'s of the realisations of headline inflation in July 2013 (for the interpretation of such *pit*'s in relation to forecast uncertainty see Rossi and Sekhposyan, 2014). They describe the probability of observing values of the random variable not greater than the observed headline inflation.

Table 8: Probabilities of hitting inflation target bands, headline inflation in July 2013 and its *p*-values

	Target interval, %	Infl. in July 2013	WSN		TPN		GBSN	
			(1)	(2)	(1)	(2)	(1)	(2)
Poland	1.5 - 3.5	1.1	0.60	0.25	0.59	0.24	0.63	0.18
Russia	5 - 7	6.5	0.43	0.38	0.79	0.56	0.43	0.26
Ukraine	4 - 6	0.0	0.07	0.34	0.40	0.21	0.00	0.32
U.S.A.	1 - 3	2.0	0.60	0.56	0.62	0.64	0.68	0.56

Legend: (1): probabilities of inflation being within the target in June 2013 according to the particular distribution;
(2): *pit*'s of observed headline annual inflation in June 2013 according to the particular distribution.

As expected, Table 1 does not show much difference between the estimated probabilities of hitting inflation target, especially for Poland and U.S.A., where all examined skew normal distributions fit well. More substantial differences can be noticed for Russia, where the results for TPN, which is the only distribution for which the null hypothesis of the uniformity of *pit*'s is not rejected, shows distinctively different probabilities of hitting inflation target and *pit* for the observed inflation. A positive conclusion which can be drawn here is that more than one skew normal distribution might fit well to the data, and it might not matter much which one is used. However, only the parameters of the estimated WSN distribution can be sensibly interpreted in the context of monetary policy.

6. CONCLUSIONS

The general message from this paper is somewhat pessimistic. We showed that it might be difficult to tell one skew normal distribution from another on the basis of the best fit, especially if the sample size is not very large. As the number of potential skew normal candidates for fitting to data is substantial (especially in the light of the fact that there are other propositions in the literature not considered in this paper) it seems to be sensible to decide on the type of distribution not on the basis of the best fit but rather on the basis of interpretation of its parameters, especially if there is not much difference in the closeness of the fit of the competing distributions. For countries conducting consistent and reasonably tight monetary policy, the weighted skew normal distribution, WSN, seems to be a sensible choice.

It is worth noting that the difficulty in deciding on the type of skew normal distribution is deepened by the fact that there are no operational statistics developed for testing the degree of disparities between distance measures (or other characteristics) of these distributions. The bootstrap procedure used in this paper suggests a way for further investigation.

References

Andrade, P. and H. Le Bihan (2013), ‘Inattentive professional forecasters’ *Journal of Monetary Economics* **60**, 967-982.

Azzalini, A., (1985), ‘A class of distribution which includes the normal ones’, *Scandinavian Journal of Statistics* **12**, 171-178.

Azzalini, A. (1986), ‘Further results on a class of distributions which includes the normal ones’, *Statistica*, **46**, pp. 199-208.

Azzalini, A. and A. Capitanio (1999), ‘Statistical applications of the multivariate skew normal distributions’, *Journal of Royal Statistical Society Series B* **61**, 579-602.

Basu, A., H. Shioya and C. Park (2011), *Statistical inference: the minimum distance approach*, CRC Press.

Berkowitz, J. (2001), ‘Testing density forecasts, with applications to risk management’ *Journal of Business and Economic Statistics* **19**, 465-474.

Bowles, C., R. Fritz, V. Genre, G. Kenny, A. Meyer and T. Rautanen (2007), ‘The ECB Survey of Professional Forecasters: a review after eight years’ experience’, Occasional paper No 59, European Central Bank.

Castro, L. M, E. San Martín and R. B. Arellano-Valle (2008), ‘A note on the identification of skewed-normal distributions’, Department of Statistics, Pontificia Universitad de Conceptión, Chile.

Charemza, W., Z. Fan, S. Makarova and Y.Wang (2012), ‘Simulated minimum distance estimators in macroeconomics’, paper presented at the conference *Computational and Financial Econometrics*, Oviedo.

Charemza, W., C. Díaz Vela and S. Makarova (2014), ‘Term structure of inflation forecast uncertainties and skew normal distributions’, University of Leicester, Department of Economics, Working Paper No. 14/01.

Clements, M. P. (2004), ‘Evaluating the Bank of England density forecasts of inflation’, *The Economic Journal* **114**, 844-866.

Clements, M.P. (2014), ‘Forecast uncertainty-ex ante and ex post: U.S. inflation and output growth’, *Journal of Business and Economic Statistics* **32**, 206-216.

Cressie, N. and T.R.C. Read (1984), ‘Multinomial goodness-of-fit tests’, *Journal of Royal Statistical Society Series B*, **46**, 440-464.

Diebold, F.X., T.A. Gunther and A. S. Tay (1998), ‘Evaluating density forecasts: with applications to financial risk management’, *International Economic Review* **39**, 863–83

Dominicy, Y. and D. Veredas (2013), ‘The method of simulated quantiles’, *Journal of Econometrics* **172**, 233-247.

Dowd, K. (2008), ‘The Swedish inflation fan charts: an evaluation of the Riksbank’s inflation density forecasts’, *The Journal of Risk Model Validation* **2**, 73-87.

Dowd, K. (2007), ‘Backtesting the RPIX inflation fan charts’, CRIS Discussion Paper Series, 2007.V

Égert, B. and R. MacDonald (2008), 'Monetary transmission mechanism in Central and Eastern Europe: surveying the surveyable', OECD Economics Department Working Paper No. 654.

Franceschini, C. and N. Loperfido (2014), 'Testing for normality when the sampled distribution is extended skew-normal', in (M. Corazza and C. Pizzi, eds.): *Mathematical and statistical methods for actuarial sciences and finance*, Springer.

Fujisawa, H. and T. Abe (21012), 'A family of skew-unimodal distributions with mode-invariance through transformation of scale', paper presented at the CFE 2012 conference, Oviedo.

Galbraith, J.W. and S. van Norden (2012), 'Assessing gross domestic product and inflation probability forecasts derived from Bank of England fan charts', *Journal of the Royal Statistical Society, Series A* **175**, 713-727.

Gómez, V. and A. Maravall (1998), 'Automatic modelling methods for univariate series', Banco de España-Servicio de Estudios. Documento de trabajo no. 9808.

Greco, L. (2011), 'Minimum Hellinger distance based inference for scalar skew-normal and skew-t distributions', *Test* **20**, 120-137.

Hall, S.G. and J. Mitchell (2007), 'Combining density forecasts', *International Journal of Forecasting* **23**, **1-13.**

Hasanalipour, P., and M. Sharafi (2012), 'A new generalized Balakrishnan skew-normal distribution', *Statistical Papers* **53**, 219-228.

John, S. (1982), 'The three parameter two-piece normal family of distributions and its fitting', *Communications in Statistics: Theory and Methods* **14**, 235-245.

Kimber, A.C. (1985), 'Methods for the two-piece normal distribution', *Communications in Statistics: Theory and Methods* **14**, 235-245.

Kowalczyk, H. (2013), 'Inflation fan charts and different dimensions of uncertainty', National Bank of Poland Working Paper No. 157.

Lahiri, K., H. Peng and X. Sheng (2014), 'Measuring uncertainty of a combined forecast', presented at the workshop 'Uncertainty and Economic Forecasting', UCL, London.

Lahiri, K. and X. Sheng (2010), 'Measuring forecast uncertainty by disagreement: the missing link', *Journal of Applied Econometrics* **25**, 514-538.

Mameli, V. and M. Musio (2013), 'A generalization of the skew-normal distribution: the beta skew-normal', *Communications in Statistics – Theory and Methods* **42**, 2229-2244.

Monti, A.C. (2003), 'A note on estimation of the skew normal and the skew exponential power distributions', *METRON – International Journal of Statistics* **20**, 205-219.

Nakatsuma, T. (2003), 'Bayesian analysis of two-piece normal regression models', presented at Joint Statistical Meeting, San Francisco.

Pewsey, A. (2000), 'Problems of inference for Azzalini's skew-normal distribution', *Journal of Applied Statistics* **27**, 859-870.

Rossi, B. and T. Sekhposyan (2014), 'Alternative test for correct specification of conditional predictive densities', paper presented at the *Seminar on Economic Modelling and Forecasting*, Warwick Business School.

Sartori, N. (2006), 'Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions', *Journal of Statistical Planning and Inference* **136**, 4259-4275.

Sharafi, M. and J. Behboodian (2008), 'The Balakrishnan skew-normal density', *Statistical Papers* **49**, 769-778.

Sikorski, K. and F. Stenger (1984), 'Optimal quadratures in H_p spaces', ACM Transactions on Mathematical Software 10, 140-151.

Sun, J., Chernick, M. and A. LaBudde (2011), 'A bootstrap test for comparing two variances: simulation of size and power in small samples', *Journal of Biopharmaceutical Statistics* **21**, 1079-1093.

Taylor, A.M.R. (2003), 'Robust stationarity tests in seasonal time series processes', *Journal of Business and Economic Statistics* **21**, 156-163.

Vdovichenko, A. and V. Voronina (2009), 'Monetary policy rules and their application in Russia', EERC Working Paper No 04/09, Economic Research Network, Russia and CIS.

Wallis, K.F. (1999), 'Asymmetric density forecasts of inflation and the Bank of England's fan chart', *National Institute Economic Review*, January, 106-112.

Wallis, K.F. (2004), 'An assessment of Bank of England and National Institute inflation forecasts uncertainties', *National Institute Economic Review* **189**, 64-71.

Yadegari, I., A. Gerami and M.J. Khaledi (2007), 'A generalization of the Balakrishnan skew-normal distribution', *Statistical and Probability Letters* **78**, 1165-1167.