ACTA UNIVERSITATIS AGRICULTURAE ET SILVICULTURAE MENDELIANAE BRUNENSIS

Volume LIX 44 Number 4, 2011

ON ROBUST ANALYSIS OF PAYCHECK: CASE STUDY

M. Uherek, M. Stehlík, L. Střelec

Received: February 28, 2011

Abstract

UHEREK, M., STEHLÍK, M., STŘELEC, L.: *On robust analysis of paycheck: case study.* Acta univ. agric. et silvic. Mendel. Brun., 2011, LIX, No. 4, pp. 371–378

Many statistical tests are constructed to check the validity of normal distribution. Here we propose a case study on analysis of paycheck where we employ the RT class of tests for normality firstly introduced in Střelec, Stehlík (2008). In particular such a study can be of interest for pension funds theoreticians and practitioners, which study the transitions of pension systems from one social security state to the another one. Our study illustrates some possible distributional deviations of salary residuals on a real data.

paycheck, case study, robust approach, normality testing, residuals

In many countries, very speed changes happen. However, pension systems are not so stable. Nonstability of pension systems is caused by many reasons. One reason is that they have been planned in totally different situation, both economical and social. Social security managers have a big challenge to understand these aspects and to stabilize the situation. Particularly, Slovakia has been following some radical changes in their pension system. For the calculation of the amount of old-age pension a relatively complicated formula was used, which in principle took account of the period of employment and the average income of the individual, determined as the average monthly income during five "best earnings" years over the period of 10 years prior to retirement. In determining the assessment base for the calculation of pension, only the "first" SKK 2,500 had been fully counted of the total income. From the SKK 2,500–6,000 range only one-third was counted, from SKK 6,000–10,000 range only one-tenth was counted and the monthly income exceeding SKK 10,000 was not considered at all. The pension calculated on this basis was subsequently adjusted upon its award by a coefficient that reflected the growth in wages after 1989 and the indexation of pensions introduced from 1991. This reduction of the assessment base was the main tool of income redistribution in the construction of pre-reform pension system; hence the people with lower income could expect considerably higher replacement rates than the persons with higher incomes. For more see (The pension system in the Slovak Republic). The existing pension systems exhibit fundamental imbalances which call for radical reform. Also knowledge sharing of solutions to such instabilities is not well distributed between both theoreticians and practitioners. It is a widely shared view, that current systems are unsustainable: hence the questions concerning the design of reform, together with the associated problem of identifying feasible patterns of transition from the inherited system to the reformed one (see also Kruse, Porta and Saraceno, 1997). For some theoretical considerations on normality testing regarding the 1st pillar in Slovak Republic see Stehlík, Střelec (2009).

The aim of this contribution is to illustrate how typical paycheck may have been looking like in the time of socialism. Policy learning and uncertainty assessment is very important for many eastern countries. The situation is rather clear: transitions and transition periods are very sensitive and policymakers should find a reliable sources to learn from them effects or possible outcomes. These sources can be based on meta-analysis of robust studies. The aim of this paper is to show a possible way of robust study of pay-check. Employment of RT class tests we consider to be very important to be able to scrutinize of rare models. For normality testing many tests has been developed. Recently, robust tests have been of interest, see robust Jarque-Bera test (RJB) introduced by Gel and Gastwirth (2008). Here we use RT class of tests introduced by Střelec, Stehlík (2008). The main idea of RT class is the Lehman-Bickel construction of location functional for the moments of classical Jarque-Bera test.

One issue which can be understand from this paper is deviation of residuals in salary regression from the hypothesized normality. Let us consider 1st pillar, which is typically based on pay-as-go system. The latter one depends completely on salaries in a long range. Two problems may encounter here: significant non-normality of a residuals sequence for salaries in a small sample size, or non-normality of residuals caused by long-term effect. For the first problem reader can be referred to Stehlík, Střelec (2009), the latter one is thoroughly discussed through this paper.

This paper is divided into four parts. Part 1 is introduction. In part 2 we explain the state-of-act of analysis of paycheck, dataset and used methods. Part 3 deals with the gist of this paper – this part presents results of data analysis and comparison of power of analysed tests for normality against p-location-outlier models. The final part of this paper is discussion and summary.

MATERIALS AND METHODS

Analysis of paycheck

In this section we would like to present some of our investigation we undertook in the concern of paycheck analysis. As the series the sets of paychecks of a remarkable length were chosen. We were able to receive a payroll from one person since 1960, which continued with some break until 1984 on monthly basis. This was a period of communism and state regulation in Slovakia. We obtained originals of the payrolls and so we need to input it into computer to be able to analyze it. For the practical reasons we chose only some series to be analyzed.

The data were obtained from one employee, Mr. Božik, who changed the job once – Tatravagonka to Zbrojovka. But there was no remarkable break and the parameters of both parts seemed to be the same, because most of the jobs were paid an hour wage, which had to be the same for all professions. As we later found in graphic analysis, the variance did not seem to be the same all over the series, but the breaking point seen later than the time of the job change. For this fact, it is reasonable to assume, that there was a change in a regime and not in the employer. This is though not the purpose of this case study, and so we discuss this issue only marginally.

The measures we found interesting to analyze were the following parameters:

- *ZD*: This time series covers a complex data about the working wages. The nature of this series (base for tax) combines wage for work done as well as the wage paid while on vacation and some other bonuses, such as bonus for children, insurance payments if injured. The bonuses paid one a year are generally paid in August and sometimes in November, which creates some kind of outlier in this series. The decision to choose this series as pre tax salary, was due to possible difference in tax rate over the period of 20 years. Naturally we did not evaluate the salary after tax, because there would be a high correlation between these two series and it would not be worth the effort of inputting the data.
- *UM*: In this series, named work wage, we monitor the basic amount of work done and rewarded. It represents the amount of hours spent working multiplied by the hour wage. In this manner it seems to be better, but we must not forget, that this series does not include the paid vacation. This causes also outlying values, in the months when there was vacation, there is a lower work wage than it should normally be. Despite this disadvantage we expect to find a series with a clear trend line and normally distributed residuals.
- *KV*: Finally we decided for the series of which represents the amount to be paid to the employee because it is not completely random and it is probably the most important from practical point of view. In general, this series is a kind of variation of series base for tax, but in contradiction, there are various deductions in form of social fund, or different types of insurance, which were (from long term point of view) randomly assigned, based on government decision in irregular intervals.

The natural assumption of growing salaries was thought about due to principles macroeconomics. We were hoping to be able to spot the average inflation rate in growing tendency or maybe a natural salary increase though it was questionable in that period. State directed manufacturing might have been restrictive for normal market behaviour. Generally speaking, a trend assumption was not doubt, despite unclear reason. Questions of economic interpretation are not the major issue here in this paper, but above mention points may be a good motivation for the reader.

Analysis of the data

Now we would like to present the notation of variables we used in the analysis, and introduce our findings regarding testing normality. Variables used in analysis:

- *ZD*, *UM*, *KV* – tax base, task wage and payment, respectivelly,
- r – a sign for the residual series (e.g. ZDr is time series of residual of ZD),
- m – a sign for the linear regression model in form of $Y = \alpha N + \beta$, where N is the number of month (e.g. ZDm is a linear regression model in form of $ZD = \alpha N + \beta$),

- *f* – a sign for the fitted data from linear regression (e.g. *ZDf* are the fitted data from linear regression).

At the early stage we found a small problem. There was a significant portion of data missing, due to damaged payroll. This is nothing extraordinary in series like the one we were analyzing. In this stage a question raise, about how to deal with this type of situation. In general, we need to decide if we analyze the time series on a monthly basis, or we take an average of salaries per year, which would create much shorter series. In this case we would need to skip first 3 years, because there is nothing to make an average of, since there are no data. In this case study, we will perform only monthly data.

To eliminate the missing values in the series, there are a few options. We can simply skip the data we do not know, which was about 10 percent of the data. The only negative incoming form this way of dealing with the series is that it would be harder to analyze the seasons or repeating cycles in the series. This is however not the objective of this paper.

Another alternative we were considering is to replace the missing data with mean or median or another measure of central tendency. But this may lead to a problem in linear regression, since the majority of missing data is at the beginning of the series.

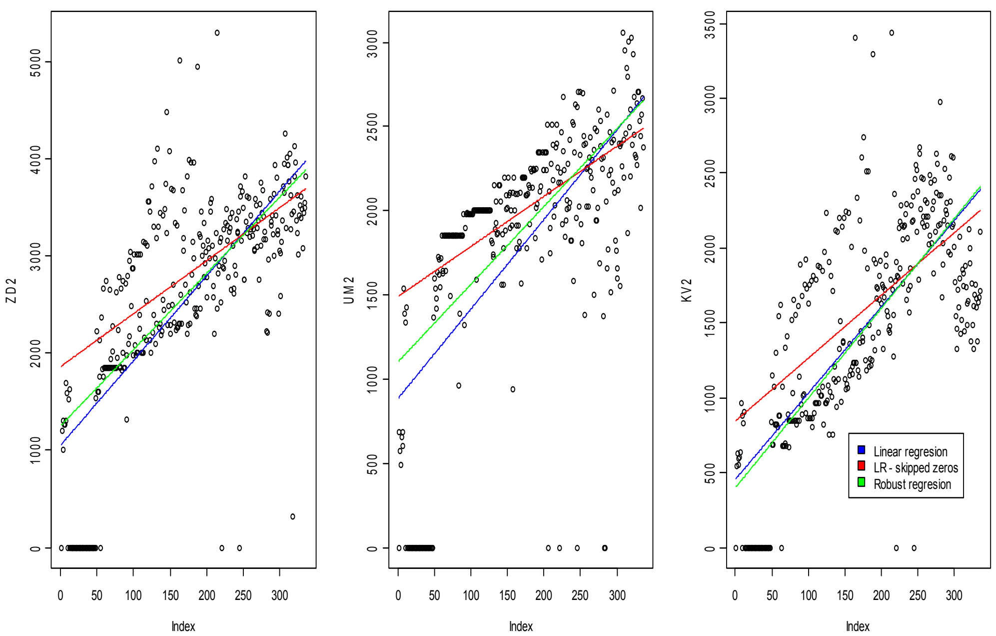
The last option we were considering was replacing the missing data with significant outliers such as zero. This would simplify the graphic analysis, because the outliers would be visible very clearly, but it would bias the series in general. But since the zeros would not be in the centre of the distribution, but on the tails, we have an option to use robust statistics to eliminate outliers.

Finally, we did not change the numbering in months, so if we skipped it, the next month has its original index, so the linear regression would apply over the same period as in the series with zeros. Our idea is to obtain residual series which we get by eliminating the trend line of the series with skipped values (the real unbiased trend line) and than perform normality tests. Since we are planning to use robust tests, we would like to analyze their power to eliminate the outlying values and provide truthful verdict about the nature of the distribution.

For testing of normality we use the general form of robust Jarque-Bera test, so called RT class, introduced by Střelec, Stehlík (2008). It was used to systematize the results from some recent studies on variants of Jarque-Bera tests and give general guidelines for appropriate small sample testing for normality. Particularly, the special cases of this class are the classical Jarque-Bera test, the Jarque-Bera-Urzua test, the robust Jarque-Bera test, the Geary's test and Uthoff's test. For comparison we also use the Anderson-Darling test (AD), the Shapiro-Wilk test (SW), the classical Jarque-Bera test (JB), the robust Jarque-Bera test (RJB), the Lilliefors-Kolmogorov-Smirnov test (LT), the skewness test (SKT) and the kurtosis test (KT) and three medcouple tests (MC1– 3) proposed by Brys *et al.* (2008).

RESULTS

Finally we decided for the first alternative mentioned above, because of the simplicity and truthfulness. We were considering the third alternative as well, but the amount of missing values is too big, to have a good result after using robust



1: *Time series and trend lines*
Source: Own calculation

procedures. For illustration we provide Figure 1 where one can easily see the difference between the trend lines. The figure shows time series, where the missing values are replaced by zeros and the lines are different trend lines regarding the series. Note that LR-skipped zeros is the regression line of the original series.

Now we would like to describe how we performed the analysis. We employed the following steps. At first we created three simple linear regression models, and so *ZDm*, *UMm*, *KVm* to be able to eliminate the trend line. We have obtained Multiple R-squared: 0.4214, Adjusted R-squared: 0.4194 F-statistic: 210.5 on 1 and 289 DF, p-value: < 2.2e-16. As one can see, the R squared values are quite low to be able to say the model is accurate, but that is caused by a large number of outliers. Similar results apply to all three models. As mentioned above, the slope of trend line can be interpreted as a combination of inflation with increases in wage based on career ladder. The effect is visible very clearly in the Figure 1 on the series *UM* (task wage) where one can observe the levels of wage increased every year. The nature of these ideas is linear, and so we decided to work with linear trend despite poor quality of regression. We were aiming to use also some higher degrees of polynomial and logarithmic trend lines, but the resulting R squared value was not increasing significantly, but we were loosing the ability to interpret the results.

Model may be more truthful if we use robust linear regression, but the amount of zero outliers seems to be so high, that even a robust regression is not able to eliminate the effect completely. Also when looking at Figure 1 we can notice, that for series *KVr*, the zeros are still within the normal scope of data, because robust regression does not skip them, but it includes them, which causes the slope to be even steeper. This is a first sign, that the zeros bias the data, and we should be more prudent when presenting a final judgment over the series.

We performed also robust linear regression for the first set of residuals (without zeros), but there was almost no effect, which means, that there is not a big number of outliers in that set. We performed as well some further analysis of correlation among the sets, identification of outliers through boxplot diagram, we tested the hypothesis of the equal parameters, and we were observing some other relationships among the series.

We tried to identify if the distributions are heavy-tailed by using likelihood ratio test statistic from Stehlík, Potocký, Waldl and Fabián (2008). We obtained empirical p-value for this test. The result was that the empirical p-value was one for all series, so they mimic to be heavy-tailed. This fact does not seem natural, because we do not expect Mr. Božik to earn an incredibly large amount in some months, despite some signs of heavy-tails in Figure 1. This heavy tails mimics is caused by the change point in the trend of regression caused by change of job.

Our assumption about the distribution is, that the residuals are normally distributed at least for series of *UM* type. This assumption rises from the nature of the data. They clearly have a growing trend tendency, and the impact on this type is only from the vacation taken in random time point.

Histograms and boxplots of residual of analysed time series are presented in Figure 2 and Figure 3, respectively.

The main objective of this section is to state a final verdict on the behaviour of the series as well as to compare the results of the standard and robust normality tests. We are going to compare the results of the tests to the original set of data. We will be able to see how the outliers influence the normality hypothesis, and we will see if the robust tests will exclude the outliers identified by boxplot analysis.

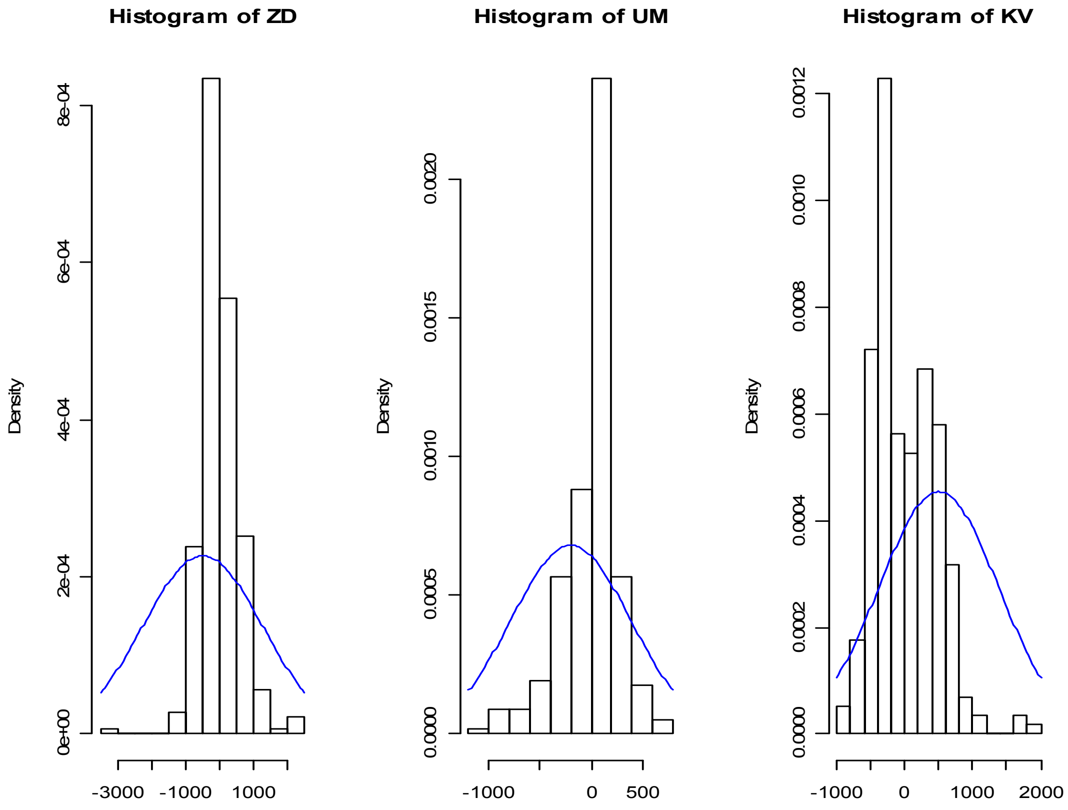
Table I presents the results of selected classical and robust tests of normality. For this purpose we use selected classical tests of normality (the classical Jarque-Bera test, the Anderson-Darling test, the Lilliefors test, the Shapiro-Wilk test, the skewness test and the kurtosis test) and selected robust tests of normality (the robust Jarque-Bera test, three medcouple tests and selected RT tests proposed by Střelec, Stehlík (2008).

To be able to analyze Table I, we have to remind of the nature of each type of the sets of data. The *UM* (task wage) type residuals are assumed to be normal. On the other hand the *KV* (payment) is showing unpredictable data, which have a growing tendency, but the residuals do not need to be normally distributed around the trend line, due to its nature. We tend to expect a linear combination of two or more distributions, which do not perform as normal after all. The *ZD* (tax base) is somewhere in the middle, because it is basically a combination of the earlier mentioned series and so the characteristic is kind of a mixture.

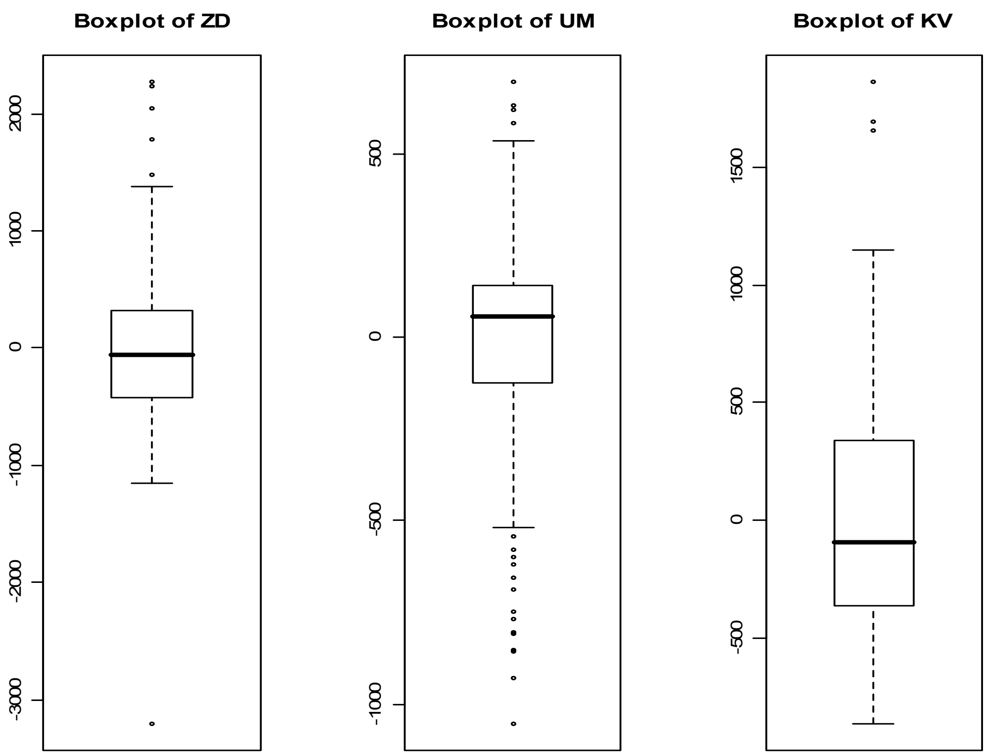
Having reviewed this information, we may start the final analysis. The LT test gives us an idea of type of distribution and helps us to decide which information obtained from JB and RJB tests are more reliable, because one of our objectives is to compare these two tests. As we can see from Table I, the p-values are all approximately zero, so we can assume all distributions not to be normal and the null hypothesis is almost always rejected. Only medcouple MC1 and MC2 show nonzero p-values.

From our findings, JB, RJB, RTJB and RTRJB tests perform very similarly. For this reason, it is worth to have a look at the reason of rejecting the hypothesis. Since we are using JB, RJB, RTJB and RTRJB tests, which are using higher degree of standard moments, we are able to decompose the test criteria into measure of skewness and kurtosis, and have a look at the type of deviation. Skewness and kurtosis measures are presented in Table II.

As we can see from Table II the main problem for *ZDr* residual series is kurtosis and for *KVr* residual series is the main problem skewness. On the other



2: *Histograms of residuals of analysed time series*
Source: Own calculation



3: *Boxplots of residuals of analysed time series*
Source: Own calculation

I: *The comparison of the results of the tests*

	ZDr		UMr		KVr	
	statistic	*p*-value	statistic	*p*-value	statistic	*p*-value
JB	275.11	0.000	86.73	0.000	38.12	0.000
AD	4.00	0.000	6.61	0.000	4.23	0.000
LT	0.11	0.000	0.14	0.000	0.12	0.000
RJB	251.49	0.000	212.15	0.000	24.90	0.001
SW	0.92	0.000	0.93	0.000	0.95	0.000
SKTst	2.03	0.041	−6.58	0.000	5.40	0.000
KTst	16.46	0.000	6.59	0.000	3.00	0.007
MC1	2.09	0.146	17.33	0.000	12.42	0.000
MC2	34.46	0.000	4.72	0.092	6.03	0.048
MC3	41.05	0.000	20.73	0.000	19.42	0.000
RTJB9	276.8	0.000	83.2	0.000	40.3	0.000
RTJB39s=r=1	294.0	0.000	75.7	0.000	25.5	0.001
RTJB39s=r=5	299.0	0.000	81.1	0.000	29.8	0.000
RTJB42s=r=1	55.8	0.000	63.1	0.000	17.3	0.000
RTJB42s=r=5	58.6	0.000	65.6	0.000	21.4	0.000
RTRJB13	256.8	0.000	155.8	0.000	26.3	0.001
RTRJB33s=r=1	13.7	0.001	101.1	0.000	24.2	0.000
RTRJB33s=r=5	16.3	0.001	107.0	0.000	23.6	0.000
RTRJB42s=r=1	52.7	0.000	99.9	0.000	13.5	0.000
RTRJB42s=r=5	130.0	0.000	245.8	0.000	17.0	0.020

Source: Own calculation

II: *The decomposition of JB and RJB test*

	ZDr		UMr		KVr	
	skewness	kurtosis	skewness	kurtosis	skewness	kurtosis
JB	4.11	271.00	43.28	43.45	29.11	9.00
RJB	6.38	245.11	92.98	119.17	24.13	0.77
RTJB9	5.81	271.00	39.73	43.45	31.33	9.00
RTJB39s=r=1	23.00	271.00	32.25	43.45	16.47	9.00
RTJB39s=r=5	27.95	271.00	37.65	43.45	20.81	9.00
RTJB42s=r=1	23.00	32.85	32.25	30.83	16.47	0.87
RTJB42s=r=5	27.95	30.63	37.65	27.95	20.81	0.62
RTRJB13	9.01	247.78	85.36	70.44	25.97	0.32
RTRJB33s=r=1	6.09	7.58	70.32	30.74	22.43	1.77
RTRJB33s=r=5	6.09	10.21	70.32	36.71	22.43	1.18
RTRJB42s=r=1	34.04	18.65	52.40	47.52	12.69	0.81
RTRJB42s=r=5	41.38	88.59	61.18	184.60	16.04	0.99

Source: Own calculation

hand, the values skewness and kurtosis of *UMr* residual series are almost equally.

Normality rejection of analysed time series could be caused by presence of outliers. For these cases, the normality tests have different power and robustness. For ilustration of this problem we consider *p*-location-outlier model *LocOut* (p, λ), where to be iid from $N(1, 0)$ and $X_{n-p+1}, \ldots, X_n$ to iid from $N(\lambda, 0)$ – see Balakrishnan (2007). In this paper we considered $n \in \{20, 100, 200\}$, $p \in \{1, 5, 10, 20\}$, and $\lambda = 5$. Consequently, Table III presents results of power of analysed normality tests against selected *p*-location-outlier models. The tests with the highest power against outlier-models are Jarque-Bera test, robust Jarque-Bera test, Shapiro-Wilk test, Anderson-Darling test, skewness test, kurtosis test and selected RT tests. On the other hand, the most robustness tests are medcouple tests, RTJB42 and RTRJB42 test, especially for small parameter p in comparison to sample size.

III: *Comparison of power of selected normality tests against p-location-outlier model*

	$n = 20$		$n = 100$				$n = 200$			
	$p = 1$	$p = 5$	$p = 1$	$p = 5$	$p = 10$	$p = 20$	$p = 1$	$p = 5$	$p = 10$	$p = 20$
JB	0.870	0.105	0.880	1.000	1.000	1.000	0.845	1.000	1.000	1.000
AD	0.644	0.920	0.348	1.000	1.000	1.000	0.220	0.999	1.000	1.000
LT	0.409	0.815	0.164	0.982	1.000	1.000	0.108	0.952	1.000	1.000
RJB	0.855	0.133	0.857	1.000	1.000	1.000	0.813	1.000	1.000	1.000
SW	0.776	0.890	0.801	1.000	1.000	1.000	0.780	1.000	1.000	1.000
SKTst	0.838	0.323	0.780	1.000	1.000	1.000	0.692	1.000	1.000	1.000
KTst	0.815	0.001	0.848	0.999	0.997	0.001	0.813	1.000	1.000	1.000
MC1	0.060	0.531	0.048	0.079	0.204	0.921	0.053	0.065	0.111	0.375
MC2	0.043	0.028	0.045	0.060	0.403	0.723	0.052	0.059	0.102	0.739
MC3	0.051	0.338	0.049	0.091	0.547	0.998	0.050	0.068	0.147	0.871
RTJB9	0.832	0.659	0.863	1.000	1.000	1.000	0.831	1.000	1.000	1.000
RTJB39s=r=1	0.854	0.602	0.871	1.000	1.000	1.000	0.831	1.000	1.000	1.000
RTJB39s=r=5	0.647	0.961	0.853	1.000	1.000	1.000	0.820	1.000	1.000	1.000
RTJB42s=r=1	0.051	0.543	0.048	1.000	1.000	1.000	0.048	1.000	1.000	1.000
RTJB42s=r=5	0.006	0.969	0.033	0.999	1.000	1.000	0.038	0.999	1.000	1.000
RTRJB13	0.837	0.540	0.841	1.000	1.000	1.000	0.793	1.000	1.000	1.000
RTRJB33s=r=1	0.868	0.025	0.815	1.000	1.000	1.000	0.698	1.000	1.000	1.000
RTRJB33s=r=5	0.860	0.143	0.817	1.000	1.000	1.000	0.702	1.000	1.000	1.000
RTRJB42s=r=1	0.073	0.367	0.074	1.000	1.000	1.000	0.068	1.000	1.000	1.000
RTRJB42s=r=5	0.081	0.617	0.127	1.000	1.000	0.916	0.140	1.000	1.000	1.000

Source: Own simulation

DISCUSSION AND SUMMARY

The aim of this paper is to illustrate robust testing for normality of salary residuals. Salary residual checks are routinely performed and sometimes robustness is not optimally applied. We have performed the analysis of the paycheck data. There may be various interpretations either for the pension system or also other fields. Our finding about the hypothesis of normality are summarized in Table I. We can state that the monthly based series of residuals are not normally distributed around their trend line. However, in practice some nonlinearities may enter salary evolutions. Then we may consider either classical nonlinear regression with precise estimation of economically interesting parameters (see Potocký, Van Ban, 1992) or we may employ a robust regression model. Both these directions are worth further investigations. The aim of this paper is also addressing robust testing of normality for residua obtained by regression analysis of payments. The results of tests are summarized in Table I. It is obvious that monthly data are not normally distributed. One reason is the outliers presence in the data set. This can be encompassed by robust tests for normality, e.g. RT tests introduced in Střelec and Stehlík (2008). These tests are useful for cases with small number of outliers present in a relatively large sample. This can be well illustrated by Table III. Thus RT class of tests is a good compromise between classical (non-robust) tests (e.g. JB, SW test) and too robust tests (e.g. medcouple tests).

Acknowledgements

The research was supported in part by AKTION Austria – Czech Republic.

Dedication

This paper is dedicated to Mr. Jozef Božik, for enabling us the data and consultations.

REFERENCES

BALAKRISHAN, N., 2007: Pernaments, order statistics, outliers, and robustness. Revista Matemática Complutense, 20(1). p. 7–107. ISSN 1139-1138.

BRYS, G., HUBERT, M., STRUYF, A., 2008: Goodness-of-fit tests based on a robust measure of skewness. Computational Statistics, Springer, 23(3), p. 429–442. ISSN 0943-4062.

GEL, Y. R., GASTWIRTH, J. L., 2008: A robust modification of the Jarque-Bera test of normality. Economics Letters 99 (1), p. 30–32, ISSN 0165-1765.

KRUSE, A., PORTA, P. L. and SARACENO, P., 1997: Pension Systems and Reforms: A Note on Transition Problems, University of Milan Bicocca. No. 02 – February 1997, [cit. 2009-09-12]. Available from: <http://dipeco.economia.unimib.it/repec/pdf/mibwpaper2.pdf >

POTOCKÝ, R., VAN BAN, T., 1992: Confidence regions in nonlinear regression models. Appl. Math. 37, p. 29–39. ISSN 0862-7940.

STEHLÍK, M., POTOCKÝ, R., WALDL, H., FABIÁN, Z., 2008: On the favorable estimation for fitting heavy tailed data, IFAS res. report Nr. 32, [cit. 2009-09-12]. Available from: http://www.ifas.jku.at/e2550/e2756/index_ger.html.

STEHLÍK M., STŘELEC L., 2009: On normality assumptions for claims in insurance, Acta univ. agric. et silvic. Mendel. Brun., LVII, No. 3, p. 141–146, ISSN 1211-8516.

STŘELEC, L., STEHLÍK, M., 2008: Some properties of robust test for normality and their applications, IFAS Research report.

The pension system in the Slovak Republic, Note for the attention of the Ageing Working Group, February 2006, Institute of Financial Policy Ministry of Finance of the Slovak Republic.

URZUA, C. M., 1996: On the correct use of omnibus tests for normality. Economics Letters, 53, p. 247–251. ISSN 0165-1765.

Address

Mgr. Michal Uherek, Zurich Insurance plc, o.z., Bratislava, Slovak Republic, RNDr. Milan Stehlík, Ph.D., Institut für angewandte Statistik, Johannes Kepler University in Linz, Freistädter Straße 315, Linz, A-4040, Austria, Ing. Luboš Střelec, Ph.D., Ústav statistiky a operačního výzkumu, Mendelova univerzita v Brně, Zemědělská 1, 613 00, Brno, Česká republika, e-mail: uherek.michal@gmail.com, Milan.Stehlik@jku.at, lubos.strelec@mendelu.cz