ETLA

ELINKEINOELÄMÄN TUTKIMUSLAITOS
THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY
Lönnrotinkatu 4 B 00120 Helsinki Finland Tel. 358-9-609 900
Telefax 358-9-601 753 World Wide Web: http://www.etla.fi/

Keskusteluaiheita - Discussion papers

No. 572

Heikki Hella

ON OUTLIERS IN TIME SERIES DATA*

* This is an enlarged version of the paper which the author presented at the OECD Meeting on Leading Indicators, Paris, October 1996.

ISSN 0781-6847

25.10.1996

HELLA, Heikki, **ON OUTLIERS IN TIME SERIES DATA.** Helsinki: ETLA, Elinkeinoelämän Tutkimuslaitos, The Research Institute of the Finnish Economy, 1996. 11 p. (Keskusteluaiheita, Discussion Papers, ISSN 0781-6847; no. 572).

ABSTRACT: Statistical routine data sets are estimated to include about 1-10 % or more of gross errors, which are the most frequent reasons for outlying observations, outliers. Outliers create many problems and confusion in statistical analysis and forecasting. The most traditional statistical methods are very sensitive to outliers. The goal of this paper is to present the basic types of outliers in a time series context, to consider the influences of outliers and how to try to prevent and diminish these negative effects on forecasts. In cyclical indicator analysis outliers are particularly troublesome as they increase the risk of false signals of direction and prospects of an economy or an industrial sector. Some recent research papers contain results that outliers seem to be associated with turning points in the business cycles; furthermore outliers tend to be clustered both within and across series; the results indicate also a dichotomy between outlier behaviour of real versus nominal time series.

Finally we consider briefly cyclical indicators and outlier robust procedures. Outliers will occur in data sets and therefore robust methods should be included in the arsenal of researchers.

KEY WORDS: Routine data sets, gross errors, outliers, seasonal adjustment, forecasting, business cycles, cyclical turning points.

HELLA, Heikki, **VIERAAT HAVAINNOT AIKASARJA-AINEISTOISSA.** Helsinki: ETLA, Elinkeinoelämän Tutkimuslaitos, The Research Institute of the Finnish Economy, 1996. 11 s. (Keskusteluaiheita, Discussion Papers, ISSN 0781-6847; no. 572).

TIIVISTELMÄ: Tässä katsauksessa tarkastellaan tilastoaineistoissa esiintyvien vieraiden havaintojen, outlierien, päätyyppejä ja outlierien aiheuttamia ongelmia tutkimusten ja ennusteiden laadinnassa. Suhdanneanalyysin ja -ennustamisen kannalta outlierit ovat ongelmallisia häiritessään perinteistä tilastollista mallinrakentamista ja lisätessään väärien ennustesignaalien riskiä. Traditionaalisten tilastomenetelmien rinnalla tarvitaan robusteja mallittamistapoja. Eräät tutkimukset viittaavat siihen, että outliereilla on taipumus rypästyä suhdannevaihtelujen käännevaiheisiin. Toinen havainto on, että hintasarjojen ja ns. reaalisarjojen outlier-ilmiö poikkeavat toisistaan. Eri tutkimuksissa on viime aikoina myös saatu viitteitä aineistojen epälineaarisuuden ja outlierien yhteydestä toisiinsa. Mitään yleispäteviä tuloksia ei kuitenkaan ole toistaiseksi onnistuttu saamaan. Outliereita tulee aina olemaan ja tietopankkien valtavaan numerotulvaan niiden on helppo kätkeytyä. Ennusteiden laatijoiden onkin tarpeellista hankkia välineitä ja valmiutta käsitellä aineiston vieraita havaintoja kyetäkseen minimoimaan niiden haitalliset vaikutukset.

AVAINSANAT: rutiinitilastot, vieraat havainnot, outlierit, kausipuhdistus, ennustaminen, suhdannevaihtelut, suhdannevaihtelujen käännevaiheet.

Yhteenveto

Katsauksessa tarkastellaan aikasarjoissa esiintyviä vieraita havaintoja, niiden rakennetta ja vaikutuksia. Rutiinitilastoissa arvioidaan olevan alasta riippuen noin 1 - 10 % havainnoista virheellisiä johtuen puhtaasti teknisluonteisista tallennus-, luokittelu- yms. tekijöistä. Tämän lisäksi tulevat erilaiset poikkeuksellisista tapahtumista, luonnonkatastrofeista, teknisistä muutoksista yms. aiheutuvat vieraat havainnot.

Ilmeisesti johtuen eri taustatekijöistä vieraille havainnoille, outliereille, ei alan kirjallisuudessa ole esitetty tarkkaa tilastomatemaattista määritelmää. Outlier-termin merkitys vaihtelee hieman eri yhteyksissä ja sitä on käytetty väljästi. Useimmiten vieras havainto määritellään havainnoksi, joka poikkeaa otoksen muista siksi paljon, että on aihetta olettaa sen olevan peräisin jostakin muusta perusjoukosta.

Outlierit aiheuttavat käytännössä monenlaisia hankaluuksia. Tässä on esimerkkinä tarkasteltu aikasarjojen kausipuhdistamiseen liittyviä ongelmia. Puhtaasti mekaaninen poikkeavien havaintojen käsittely voi jopa aiheuttaa lisäongelmia ja joskus arvokkaan informaation haaskaamista. Jos outlierit pystytään täsmentämään luotettavasti ja niille saadaan selkeä tulkinta, kausivaihtelunkin estimointi on vankemmalla pohjalla.

Aikasarja-analyysissa tavattavien outlierien perustyypit poikkeavat lineaarisen regressioanalyysin selitysmalleissa havaittavista. Toinen erottava piirre on vieraiden havaintojen rypästyminen ja sen aiheuttamat ongelmat aikasarjojen käsittelyssä. Aikasarjojen tapauksessa erityisen "pahoja" ovat ns. additiiviset outlierit, jotka "hoitamattomina" voivat heikentää ennusteiden laatua vielä myöhemmässä vaiheessa.

Vieraiden havaintojen esiintyminen suhdannesarjoissa ei ole mikään uusi asia. Eräiden viimeaikaisten tutkimusten antamat viitteet outlierien rypästymisestä suhdanteiden käännevaiheisiin ja vaikutukset sarjoista toiseen ovat tärkeitä sekä ennusteiden tekijöille että tilastotieteilijöille ja ekonometrikoille. Toinen todettu seikka on kaksijakoisuus hinta- ja reaalisarjojen outlier-tulosten välillä: eroja on havaittu mm. outlierien voimakkuudessa, ajoittumisessa ja tyypeissä. Tarvitaan luonnollisesti paljon lisää tutkimusta näiden ilmiöiden yleispätevyyden vahvistamiseksi.

Tutkijat ovat havainneet merkkejä yhteyksistä aikasarjojen epälineaarisuuden ja outlierien välillä. Aiheesta voi lähitulevaisuudessa kehittyä hyvinkin mielenkiintoinen tutkimuskohde. Tutkijoiden ja mallinrakentajien on jatkossa joka tapauksessa tarpeellista varustautua myös keskeisillä robusteilla analyysivälineillä.

ON OUTLIERS IN TIME SERIES DATA

1. Introduction

Modern, large data banks include thousands of routine time series, panel data and cross-sectional data sets. The high quality of data material is the goal of every research institute or organization. However, statistical routine data are estimated to include about 1-10 % or more gross errors which are the most frequent reasons for outlying observations, outliers (Hampel et al., 1986, pp. 25-28). Outliers have long been recognized as a problem in time series analysis. In leading indicator analysis outliers can be particularly troublesome as they increase the risk of false signals of direction and prospects of an economy or an industrial sector.

The term outlier is mostly used rather informally. In the statistical literature there are different definitions; we can say that an outlier is an observation (or a subset of observations) which appears to be inconsistent with the remainder of that set of data.

Most of the standard statistical methods are not able to protect against outliers. It is known that only one outlying (aberrant) observation can ruin the sample autocorrelation (cross-correlation) function estimates and the least squares estimate. The spectrum estimator is non-robust too. The most known robust statistics in practice seem to be the median and the trimmed mean.

The goal of this short paper is to present the basic types of outliers[1] in time series context and to consider briefly some results of outlier analysis on business cycles estimates. The results in the literature show that the outlier analysis is important especially in financial data sets. This means more work, but e.g. in forecasting even a small decrease in the risk of false signals is worth the trouble. There are many computer algorithms and programs available for robust basic statistics and linear regression. In case of time series analysis the supply of robust packages has been quite scant, but is increasing.

2. An example: seasonal adjustment procedures and outliers

Statisticians know well that in the seasonal adjustment procedure outliers (e.g. strike effects) in time series have been a confusing factor in estimating seasonal components. In general the

---

[1] We consider here the outliers in quantitative data. The outliers in qualitative data are conceptually different from outliers in quantitative series: the values themselves in qualitative case can not be extreme, since there are a limited (determined) number of categories in questionaires, but we can speak of outliers if frequencies (in different categories) are deviant, very low or high. In general, outliers can occur more regularly in some data sets than in others. For example, there seem to be clear difference between the macroeconomic time series and the cross-sectional data sets; in cross-sectional data outliers are more likely to occur (see, e.g. Guerard, 1988 and Virtanen & Yli-Olli, 1989).

seasonal adjusting procedures include mechanical algorithms for cleaning (smoothing) the extreme values of the irregular component based on some criteria as a threshold.

There are some arguments against the mechanical treatment of outliers[2]:

1. in removing extreme values it is possible that we may lose invaluable information; as it is known the outliers of a time series are not necessarily the extreme values and vice versa

2. if e.g. the normality test indicates that there are aberrant observations in data, we should use robust estimators

3. the proper interpretation and modelling of outliers can be crucial for the careful identification of other components of time series; this is important especially in the case of leading indicators and synthetic index; e. g. the adjusting factor of strike influences is often exponentially dampening and needs some modelling (see model 4 in Appendix)

4. it is a very difficult situation if the observation at the end of a series is an outlier (or there is a patch of outliers), because we are then not able to identify the type of outlier. If we have used robust tools to identify and estimate the previous outliers, we can make the forecast with more confidence (Hillmer, 1984)

5. if we must make the time series stationary by differencing, the outlier problem gets worse, because after differencing the number of outliers is increased and their type is more difficult to identify.

3. Outliers in time series data

There is a clear theoretical distinction between the outlier analysis of traditional linear regression analysis and time series analysis.[3] The basic distinction is the autocovariance structure of time series. In the time series context we have different types of outliers.

The initial work on outliers in univariate, stationary time series was done by Fox (1972). He considered the two basic forms:

a) additive outlier (**AO**) and (Fox calls it: Type I outlier)
b) innovational outlier (**IO**) (Fox: Type II outlier)

---

[2] Balke & Fomby (1994, p. 185) remark that the use of seasonally adjusted data probably makes it even more difficult to uncover outliers.

[3] The history of outliers and robustness is very long. The discussion about the appropriateness of rejection of outliers goes back at least as far as Daniell Bernoulli (1777) (see, Hampel et al., 1986). The interest in this part of statistics is again increasing; the ECAS (European Courses in Advanced Statistics) course "Robustness in Statistics, Theory and Applications" was held in Germany, October 1989. The robust theory of linear regression analysis is advanced and there is a comprehensive literature of the robust methods and applications, while in time series context robust theory and applications are in an intensive stage of development.

These outliers have a different influence on the observations of the series. The classical gross errors such as typing or decimal point errors (i.e. recording errors) are examples of additive outliers. Additive outliers can occur as single-point (isolated) outliers or as a patch of them. A block of additive outliers can sometimes be modelled as so-called reallocation outliers (Wu et al., 1993). We add to the outlier group also the level shift, **LS,** and a temporary (or transient) change, **TC** (Tsay, 1988). The detailed mathematical form of these outlier models is presented in the Appendix.

Testing for unit roots in time series has become a common practice when modelling univariate and multivariate economic time series from which the most are known to show a nonstationary pattern. There is evidence that e.g. the additive outliers and temporary change outliers will establish the wrong indication that a time series is stationary when it is actually nonstationary. Taking account the outliers in the series can lead to nonrejection of unit root hypotheses (Lucas, 1996, p. 89).

The essential features in outlier analysis[4] are a) the simultaneous modelling of outliers and the components of a time series or b) the combined use of outliers with a so-called intervention model (Box & Tiao, 1975) and the simultaneous modelling of other components of a series. One useful area for the outlier robust tools is to calculate estimates of missing or deleted values of a time series; we can estimate the missing or deleted observation by a recursive estimation of an additive outlier (Ljung, 1993).

4. Outliers and forecasts

In routine time series forecasting little attention is usually given to outlying observations. What are the consequences of ignoring occasional outliers? How do unrecognized outliers affect the point forecasts and the prediction intervals? If we consider the case of ARIMA models[5], Ledolter (1989) has shown, that when we have the estimated model, the additive outliers affect:

(i) point forecasts only moderately, provided that the outliers occur not too close to the forecast origin; if the AO is around the forecast origin, it has a disastrous effect on the forecast performance (see also, Hillmer, 1984)

---

[4] There are two distinct approaches to handling time series outliers:

a) the accommodation approach, which means that the goal is protection against any outlier by using robust parameter estimation methods as M- and GM-estimators (Hampel et al., 1986); the outliers themselves are not relevant as the object of research.
b) the outlier detection and interpretation approach, in which the modelling of outliers e.g. for forecasting is important.
In addition, some Bayesian methods are developed to handle the problems of outliers in time series (see Barnett and Lewis, 1984).

[5] Naturally we can analyse outliers in the framework of the transfer function noise models and the structural time series models (Harvey, 1989).

(ii) increase the width of ARIMA prediction intervals, because these intervals are proportional to the estimate of the standard deviation of the innovations of the model.

The magnitude of the outliers increases the bias of parameter estimates and the innovation standard deviation (or variance) estimate. This variance estimate is quite sensitive to outliers. It is important to note, that as a consequence of the very biased parameter estimates, the fitted model can sometimes lead to so-called spurious outliers ; this phenomenon is called swamping.

The research results point out that the effects of an outlier depend on its magnitude, position in the time series and type. The situation is more difficult if we have both isolated and patchy outliers in the same series and further of different type. The research results also indicate that the innovational outliers, which are caused by external shocks, do not in general have such severe consequences as the additive outliers. IO can be in some situations problematic because the outlier is added to the innovations series of ARIMA and the effects of the outlier on the observed series is propagated to all observations following time $t = T$ according to the (polynomial) structure of ARIMA (see model 2 in Appendix); this effect is known as smearing.

The unrecognized outliers of the individual series may have distorting influences on the composite indices (or synthetic indices) and thereby the forecasts. As Hillmer (1984) points out, an additive outlier will affect not only the forecast error for the period in which it occurs but also the forecast errors of subsequent periods. The distortion can be more serious because of the interaction between outliers within a series and/or across series (see e.g. Burman & Otto, 1988 and Balke & Fomby, 1994). The best way to try to resolve this dilemma, if possible, is to seek any piece of information about the potential causes of detected possible events.

In practice, the most difficult problem seems to be the so-called **masking effect** in which the outlier covers one or many other outliers of a series so that the test can fail to detect any outliers. This can happen, if we apply the ordinary, non-robust least squares estimation method (LS) to a series, which includes outliers. The residual analysis with the least squares method does not identify all the outliers present in the data. The typical cases can be found in the share price or similar financial time series (e.g. Chatterjee & Jacques, 1994). Chatterjee and Jacques apply succesfully a very robust estimation method, reweighted least median squares (RWLMS), which is one of the high breakdown point estimation methods (Hampel et al., 1986).

5. Outliers and cyclical indicators

What should we do if we know that an outlier or a subset of outliers exists or if there are some signals of possible outliers in a series? There is no unique answer regarding the appropriate device. But, if we have in our arsenal of statistical tools some robust techniques, these with the routine explorative data analysis like the normality test, and, of course, the graphical analysis (normality paper and scatter diagram analysis) are relevant.

If the preliminary analysis indicates some contamination in the series, the following steps are essential:

(i) try to detect and identify all aberrant observations as perfectly as possible using some robust scale and location estimator in outlier modelling perhaps with a robust smoother (or filter) technique
(ii) try to find the proper interpretation for as many outliers as possible
(iii) if the points (i)-(ii) are successful, we can apply also the combination of intervention and outlier analysis.

For the points **(i)**-**(iii)** many iterative detection and estimation procedures are developed in the literature (e.g. Hillmer et al., 1983). Some of them include the joint estimation of time series model parameters and outlier effects (e.g. Tsay, 1988; Chen and Liu, 1993; TRAMO, Gomez & Maravall, 1994). One interesting iterative procedure developed by Burman & Otto (1988), who fitted sixty Census Bureau monthly time series with ARIMA models, identified additive outliers and sought their external causes. They also discussed about the testing for outlier detection, and the effects of the outlier analysis on the seasonal adjustment of time series[6].

After the outlier analysis we have the outlier-adjusted time series for use as the single leading series or as part of either a composite indicator (synthetic index) or a time series forecasting model. As Guerard et al. (1995) point out, the outlier analysis (outlier-adjusted) time series model of GNP can outperform its rivals. The criterion was the post-sample forecasting accuracy.

In practice we often encounter the point of aggregation, especially in business cycle analysis. We can separate time aggregation and aggregation of subseries. The effects of aggregations can be remarkable both in forecasting and in outlier analysis. As Balke & Fomby (1994) note, time aggregation tends to obscure the presence of large shocks. They found e.g. that the outlier search in the monthly series does pick up episodes missed in the quarterly series and the timing of certain outliers can change due to aggregation. In outlier analysis of the consumption series Balke & Fomby refer to the result that if consumption is disaggregated into consumption of durables, non-durables and services, most of the outliers come from consumer durables (Balke & Fomby, 1994, p. 192).

The aggregation effects can occur also in the survey data context. It should be very interesting to investigate the frequency, timing and type of outliers in subseries of large, medium-size and small companies. For classified financial panel data this kind of outlier analysis produces very interesting results (see e.g., Chatterjee & Jacques, 1994).

The other important feature is that the pattern of outliers in the real output and employment series can be different from that of the nominal price series. Balke & Fomby (1994) found this dichotomy between real output and aggregated price series of US macroeconomic time series[7]. This concerns the differences in timing, type and frequency of occurrence of outliers; e.g. the

---

[6] After the original version of this paper in June 1996, the Journal of Forecasting published in July the article of Bruce and Jurke about two robust procedures of seasonal adjustment: the X-12-ARIMA method of the US Bureau of the Census and the MING procedure (Bruce and Jurke, 1996).

[7] It is interesting to note the results of Takala and Virén (1995) about the nonlinearity of the Finnish macroeconomic time series data (1920-1994). They found certain differences between nominal and real variables with respect to nonlinear behaviour. They did not apply any outlier modelling or other outlier analysis.

frequency of occurrence is in general higher in price series than in real output series. For leading indicator analysis this result is already very important; however, further research is needed.

Clustering of outliers within and across series and associating with turning points of business cycles[8] are difficult to handle also theoretically, as Balke & Fomby indicate, and for business cycle forecasting it is really a challenge. The statistical multivariate outlier analysis should be useful, but it needs much development and applications (see e.g., Lucas, 1996).

Some researchers have during last years noted that there are indications of connections between the outliers and non-linearity[9]. In business cycles this is especially interesting. Balke & Fomby made some tests and found that after fitting the outlier model and controlling for the effects of the outliers, the evidence of non-linearity was in many series substantially weaker (Balke & Fomby, 1994, p. 195). As they comment much further investigation is, naturally, needed.

6. The cyclical indicators and outliers

The robust outlier analysis is a modern set of complementary tools for statistician's arsenal of procedures. In the framework of large number of various indicator series it is reasonable for safety to have available both standard and robust statistical tools. In uncertain cases it is important to use both kinds of tools and compare the results.

An arsenal of robust procedures (both ordinary and sophisticated) may contain useful tools when we control the quality of the present indicators and select the new ones. In fact, at least some of these robust procedures should be a part of the system of quality control of statistical data (see also Hawkins, 1980).

For leading indicator analysis and for cyclical indicators on the whole the development of the outlier robust methods and procedures are important e.g. for the following reasons:

* there are and will exist both isolated and patchy outliers in time series data sets
* outliers tend to be associated with turning points in the business cycles
* the mechanical procedures of outlier handling in the routine seasonal adjusting programs should be replaced by robust outlier and intervention modelling
* the large routine data banks are quite exposed to outliers (outliers can easily be "hidden" in huge amount of figures)
* the structure of indicators is changing, more short-, real-time data, more survey data etc.
* there are risks of errors in data analysis: a) in combining of data for an indicator from several sources or b) in using the data of different structure e.g. of the different parts of a very long time series

---

[8] The study of Virtanen and Yli-Olli (1989) based on the cross-sectional data of the Finnish firms (41 firms, yearly data) produced the results which indicate some kind of association of outliers with the business cycles.

[9] See e.g. Harvey (1989, p. 333), Petruccelli (1990, p. 35), Tsay (1991, p. 437) and Teräsvirta and Anderson (1992, p. 135).

* the behaviour of households and companies has changed essentially during last years and it is hard for us to know what kind of change is going on
* the consequences of false signals in forecasting are more severe nowadays than earlier
* the search and analysis of outliers should be seen as an integral part of modern time series analysis and forecasting.

REFERENCES

Balke, N.S. and Fomby, T.B. (1994): Large shocks, small shocks, and economic fluctuations: outliers in macroeconomic time series. Journal of Applied Econometrics, 9, 181-200.

Barnett, V. and Lewis, T. (1984): Outliers in Statistical Data, Second Edition. John Wiley & Sons, New York.

Box, G.E.P. and Tiao, G.C. (1975): Intervention analysis with applications to environmental and economic problems. JASA, 70, 70-79.

Bruce, A.G. and Jurke, S.R. (1996): Non-Gaussian seasonal adjustment: X-12-ARIMA versus robust structural models. Journal of Forecasting, 15, 4, 305-328.

Burman, J.P. and Otto, M.C. (1988): Outliers in time series. Statistical Research Division Report Series: CENSUS/SRD/RR-88/14. Washington DC: Bureau of Census.

Chatterjee, S. and Jacques, W.E. (1994): An outlier-resistant approach to risk estimation. Financial Analysts Journal, September-October, 69-75.

Chen, C. and Liu, L.-M. (1993): Forecasting time series with outliers. Journal of Forecasting, 12, 13-35.

Fox, A.J. (1972): Outliers in time series. JRSS, B, 34, 3, 350-363.

Gómez, V. and Maravall, A. (1994): Program **TRAMO** "**T**ime Series **R**egression with **A**RIMA Noise, **M**issing Observatios, and **O**utliers". European University Institute, Badia Fiesolana, Italy.

Guerard, J.B. Jr (1988): Stationarity and outliers in annual accounting earnings. Economics Letters, 26, 369-373.

Guerard, J.B. Jr, Blin, J., Bender, S., Stone, B.K. and Wolfe, J.T. (1995): Outlier time-series models and analysts' forecasting of GNP and corporate earnings per share. Applied Financial Economics, 5, 113-119.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986): Robust Statistics The Approach Based on Influence Functions. John Wiley & Sons, New York.

Harvey, A.C. (1989): Forecasting, structural time series models and the Kalman filter. Cambridge University Press, Cambridge.

Hawkins, D.M. (1980): Identification of Outliers. Chapman and Hall, London.

Hillmer, S.C. (1984): Monitoring and adjusting forecasts in the presence of additive outliers. Journal of Forecasting, 3, 205-215.

Hillmer, S.C., Bell, W.R. and Tiao, G.C. (1983): Modeling considerations in the seasonal adjustment of economic time series. In Applied time series analysis of economic data, A. Zellner (ed.). Washington, DC: US Bureau of the Census, pp. 74-100.

Ledolter, J. (1989): The effect of additive outliers on the forecasts from ARIMA models. International Journal of Forecasting, 5, 231-240.

Ljung, G.M. (1993): On outlier detection in time series. JRRS, B, 55, 2, 559-567.

Lucas, A. (1996): Outlier Robust Unit Root Analysis. Ph.D. thesis. Tinbergen Institute Research Series no. 111, Amsterdam.

Mills, T.C. (1990): Time series techniques for economists. Cambridge University Press, Cambridge, UK.

Petruccelli, J.D. (1990): A comparison of tests for SETAR-type non-linearity in time series. Journal of Forecasting, 9, 25-36.

Takala, K. and Virén, M. (1995): Testing nonlinear dynamics, long memory and chaotic behaviour with macroeconomic data. Bank of Finland, Discussion Papers 9/95.

Teräsvirta, T. and Anderson, H.M. (1992): Characterizing nonlinearities in business cycles using smooth transition autoregressive models. Journal of Applied Econometrics, 7, S119-S136.

Tsay, R.S. (1988): Outliers, level shifts and variance changes in time series. Journal of Forecasting, 7, 1-20.

Tsay, R.S. (1991): Detecting and modeling nonlinearity in univariate time series analysis. Statistica Sinica, 1, 431-451.

Virtanen, I. & Yli-Olli, P. (1989): Cross-sectional properties and time-series persistence of financial ratio distributions; empirical evidence with Finnish data. European Institute for Advanced Studies in Management, Working Paper 89-04.

Wu, L.S., Hosking, J.R. and Ravishanker, N. (1993): Reallocation outliers in time series. Applied Statistics, 42, 2, 301-313.

JASA = Journal of American Statistical Association
JRRS = Journal of Royal Statistical Society.

APPENDIX

Outlier models in time series

We can have either seasonal or nonseasonal ARMA processes. To simplify the presentation only the nonseasonal case without a constant term is presented. Let $\{Y_t\}$ be a time series following an ARMA process

$$Y_t = \frac{\theta(B)}{\alpha(B)\phi(B)} a_t \qquad t = 1,\dots, n,$$

where n is the number of observations for the series; the polynomials $\theta(B)$ and $\phi(B)$ have all roots outside the unit circle; all roots of $\alpha(B)$ are on the unit circle; a_t is the innovation process $\approx N(0,\sigma^2)$. B is the backshift operator.

1. Additive outlier model (AO)

An additive outlier is an event that affects a series for one time period only; its effects are independent of the ARIMA model. If we assume that an outlier occurs at time t = T, we have the AO model

$$Z_t = Y_t + o_A P_t^{(T)},$$

where $P_t^{(T)}$ is a pulse function (i.e. $P_t^{(T)} = 1$ when t = T and is 0 otherwise) and o_A is the magnitude of the additive outlier. Note that Z_t is the observed, contaminated time series and Y_t is the uncontaminated core prosess (here unobserved at time t = T).

When a block (a patch) of unusually high and low values of observations occurs in such a way that the sum of the observations within the block is the same as might have been expected for a regular series, we call the observations reallocation outliers (Wu et al., 1993). A typical situation occurs in a series of sales: the period of a sales promotion and its aftermath.The constraint of the sum is what distinguishes this kind of patch from the patch of ordinary additive outliers.

One special use of the AO models in practice is the calculation the estimates of missing or deleted observations of a time series. There are developed several algorithms (see, e.g. Ljung, 1993).

2. Innovational outlier model (IO)

An innovational outlier is an event whose effect is propagated according to the structure of the ARMA model of Y_t. The IO model is then

$$Z_t = Y_t + \alpha(B)^{-1} \frac{\theta(B)}{\phi(B)} o_I P_t^{(T)}, \quad \text{or} \quad Z_t = \frac{\theta(B)}{\alpha(B)\phi(B)} (a_t + o_I P_t^{(T)})$$

where $P_t^{(T)}$ is a pulse function and o_I is the magnitude of the innovational outlier at time t = T.

3. Level shift model (LS)

$$Z_t = Y_t + \frac{1}{1-B} o_L P_t^{(T)},$$

where o_L is the permanent level shift being initiated at time t = T.

4. Temporary change model (TC)

$$Z_t = Y_t + \frac{1}{1-\delta B} o_C P_t^{(T)} \quad 0 < \delta < 1,$$

where o_C is the magnitude (initial impact) of the outlier at time t = T. The effect of this outlier decays exponentially according to the dampening factor δ. If we know the timing of the outliers, by modifying the model 4 we can easily estimate an intervention model (see, Box and Tiao, 1975; Mills, 1990).

For the models (1-4) there are developed the likelihood ratio test statistics for testing the existence of an outlier or level change (see, e.g. Tsay, 1988).

E L I N K E I N O E L Ä M Ä N   T U T K I M U S L A I T O S (ETLA)
THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY
LÖNNROTINKATU 4 B, FIN-00120 HELSINKI

Puh./Tel. (09) 609 900
Int. 358-9-609 900

Telefax (09) 601753
Int. 358-9-601 753

KESKUSTELUAIHEITA - DISCUSSION PAPERS ISSN 0781-6847

No 544 REIJA LILJA, Career Mobility in Finnish Industry. 14.12.1995. 22 p.

No 545 JUKKA LASSILA - TARMO VALKONEN, Policy Credibility in Numerical Overlapping Generations Models. 28.12.1995. 28 p.

No 546 EIJA KAUPPI - JUKKA LASSILA - TIMO TERÄSVIRTA, Short-Term Forecasting of Industrial Production with Business Survey Data: Experience from Finland's Great Depression. 26.01.1996. 20 p.

No 547 JIAN-GUANG SHEN, FDI, Knowledge Spillover and Economic Growth in East Asia. 31.01.1996. 70 p.

No 548 MAARIT SÄYNEVIRTA - PEKKA YLÄ-ANTTILA, Integraation ja yritysten kansainvälistyminen - vaikutuksia kotimaan talouteen. 01.02.1996. 58 s.

No 549 LAURA PAIJA - PEKKA YLÄ-ANTTILA, The Impact of Structure and Competition on Employment in the Telecommunications Cluster Case Finland. 12.02.1996. 34 p.

No 550 PENTTI VARTIA - PEKKA YLÄ-ANTTILA, Technology Policy and Industrial Clusters in a Small Open Economy - The Case of Finland. 16.02.1996. 15 p.

No 551 PONTUS BRAUNERHJELM - PER HEUM - PEKKA YLÄ-ANTTILA, Internationalization of Industrial Firms. Implications for growth and industrial structure in the Nordic countries. 16.02.1996. 33 p.

No 552 REIJO MANKINEN, Alkoholiveron alentamisen kansantaloudellisia vaikutuksia. 26.02.1996. 42 s.

No 553 RITA ASPLUND, Koulutus, työura ja palkkaerot. 22.03.1996. 13 s.

No 554 MARIANNE PAASI, The Absorptive Capacities of Estonian Firms. - Can a Technology-based Industrial Strategy Succeed? 22.03.1996. 17 p.

No 555 HANNU HERNESNIEMI, Barriers to Economic Cooperation of Baltic Rim Countries. 10.04.1996. 46 p.

No 556 ANNICK LARUELLE - MIKA WIDGRÉN, Is the Allocation of Voting Power among the EU States Fair? 17.04.1996. 19 p.

No 557 JARI HYVÄRINEN, A Survey of Corporate Governance - Which Model for Transition Countries? 13.05.1996. 32 p.

No 558 PASI KUOPPAMÄKI, Joint Implementation ilmastopolitiikan välineenä: Suomi ja lähialueiden kasvihuonekaasujen päästöjen rajoittaminen. 12.06.1996. 35 s.

No 559 MIKA PAJARINEN, Työnantajan kansaneläkemaksun porrastuksen toimivuus vuosina 1993 ja 1994. 20.6.1996. 34 s.

No 560 OKKO-PEKKA SALMIMIES, EU:n pakolais- ja maahanmuuttoasioita koskevan yhteistyön kehittäminen ja taloudellinen ulottuvuus. 10.07.1996. 32 s.

No 561 AIJA LEIPONEN, Education, Tenure and Innovation in Manufacturing firms. 16.08.1996. 26 p.

No 562 AIJA LEIPONEN, Education and Innovative Capabilities. 16.08.1996. 20 p.

No 563 AIJA LEIPONEN, Competences, Innovation and Profitability of Firms. 16.08.1996. 16 p.

No 564 JYRKI RUUTU, Suomen valuuttakurssijärjestelmä osana eurooppalaista järjestelmää. Historiallinen katsaus keskiajalta autonomian ajan loppuun. 29.08.1996. 22 s.

No 565 HEIKKI PALM, Eläkeuudistuksen vaikutukset Suomen kansantalouden numeerisessa limittäisten sukupolvien mallissa. 02.09.1996. 34 s.

No 566 JYRKI ALI-YRKKÖ, Teknologiaintensiivisten yritysten kansainvälistyminen - vaikutuksia Suomen talouteen. 05.09.1996. 53 s.

No 567 VENLA SIPILÄ, Suomen ja Venäjän välinen kauppa 1986-1995. 06.09.1996. 100 s.

No 568 EUGEN KOEV, Palkkadiskriminaatio teollisuuden toimihenkilöillä. 11.09.1996. 73 s.

No 569 MARIANNE PAASI, The Inherited and Emerging Absorptive Capacities of Firms - Results of a firms survey in the Estonian electronics industry. 24.09.1996. 26 p.

No 570 TANJA KIRJAVAINEN - HEIKKI A. LOIKKANEN, Efficiency Differences of Finnish Senior Secondary Schools: An Application of Dea and Tobit-analysis. 02.10.1996. 40 p.

No 571 ATRO MÄKILÄ, Teknologian työllisyysvaikutukset - Katsaus kirjallisuuteen. 21.10.1996. 19 s.

No 572 HEIKKI HELLA, On Outliers in Time Series Data. 25.10.1996. 11 p.

Elinkeinoelämän Tutkimuslaitoksen julkaisemat "Keskusteluaiheet" ovat raportteja alustavista tutkimustuloksista ja väliraportteja tekeillä olevista tutkimuksista. Tässä sarjassa julkaistuja monisteita on mahdollista ostaa Taloustieto Oy:stä kopiointi- ja toimituskuluja vastaavaan hintaan.

Papers in this series are reports on preliminary research results and on studies in progress. They are sold by Taloustieto Oy for a nominal fee covering copying and postage costs.

d:\ratapalo\DP-julk.sam/25.10.1996