Advances in historical time series analysis

Diebolt, Claude

Historical Social Research, Vol. 32 — 2007 — No. 3, 319-324

CLIOMETRICS

Advances in Historical Time Series Analysis

Claude Diebolt [*]

Abstract: *»Fortschritte in der Historischen Zeitreihenanalyse«*. This article summarises a new econometric technique for shock analysis in historical economics (cliometrics): the outlier methodology.

In cliometrics we constantly observe that regular shocks are superposed by irregular shocks which appear rarely (infrequent large shocks). This includes the question whether long-term economic development is caused (or not) by such extraordinary shocks such as wars, political measures and institutional changes. If this was the case, economic growth and development could probably not be explained as a systematic endogenous process but would have to be traced back to specific historical events. In view of that, this article summarises a new econometric technique for shock analysis in historical economics: the outlier methodology:[1]

> Outliers represent sudden temporary or permanent shifts in the level of a time series. There are several methods for the detection of outliers based on intervention analysis as originally proposed by Box and Tiao (1975). An often used procedure is that of Tsay (1988). This method was also used by Balke and Fomby (1994), although with some modifications. Here we will use an improved algorithm by Chen and Liu (1993), which is readily available, with slight modifications, in the computer program TRAMO developed by Gómez and Maravall (1997, 2001).[2]

---

[*] Address all communications to: Claude Diebolt, BETA/CNRS, Université Louis Pasteur de Strasbourg, Faculté des Sciences Economiques et de Gestion, 61 Avenue de la Forêt Noire, 67085 Strasbourg Cedex, France; e-mail: cdiebolt@cournot.u-strasbg.fr; URL: www.cliometrie.org.

[1] For the reader interested in cliometric applications and a complete mathematical and statistical presentation of the outlier methodology, please cf. Darné and Diebolt, 2004, 2006; Metz, 2002, 2004.

[2] Darné and Diebolt, 2004, p. 1452.

Consider a univariate time series y_t^* which can be described by the ARIMA(p, d, q) model:

$$\alpha(B)\phi(B)y_t^* = \theta(B)a_t \quad (1)$$

where B is the lag operator, a_t is a white noise process, $\alpha(B), \phi(B), \theta(B)$ are the lagged polynomials with orders *d, p, q*, respectively. The outliers can be modelled by regression polynomials as follows:

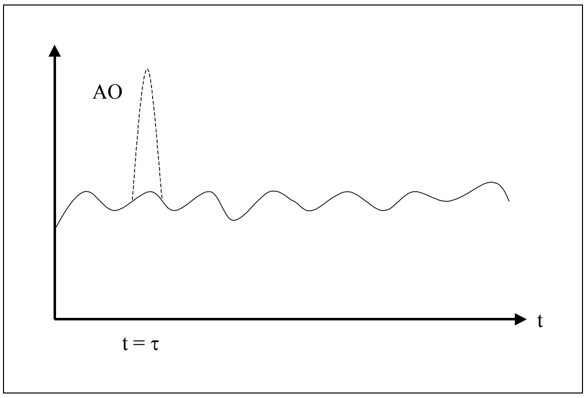
$$y_t = y_t^* + \sum_I \omega_i v_i(B) I_t(\tau) \quad (2)$$

where y_t^* is an ARIMA process, $v_i(B)$ is the polynomial characterizing the outlier occuring at time t = τ, ω_i represents its impact on the series and $I_t(\tau)$ is an indicator function with the value 1 at time t = τ and 0 otherwise.

In this paper, four main outliers are classified as:

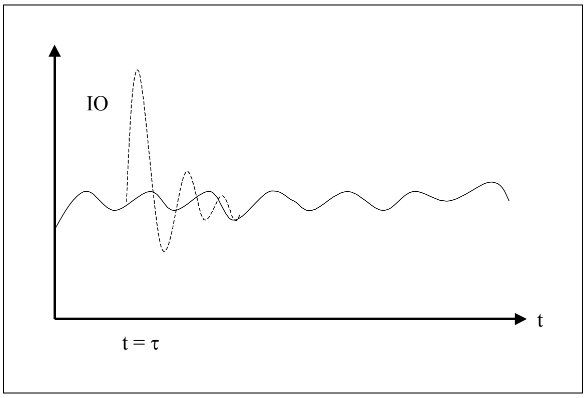
- Additive Outliers (AO) that affect only a single observation at some points in time series and not its future values. In terms of regression polynomials, this type can be modelled by setting: $v_1(B) = 1$.

Fig.1: Additive Outliers



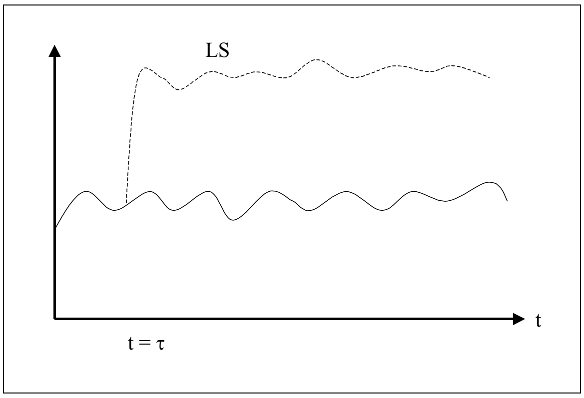
- Innovational Outliers (IO) that affect temporarily the time series with the same dynamics as an innovation. The polynomial is then $v_i(B) = \theta(B)/\phi(B)$.

Fig.2: Innovational Outliers



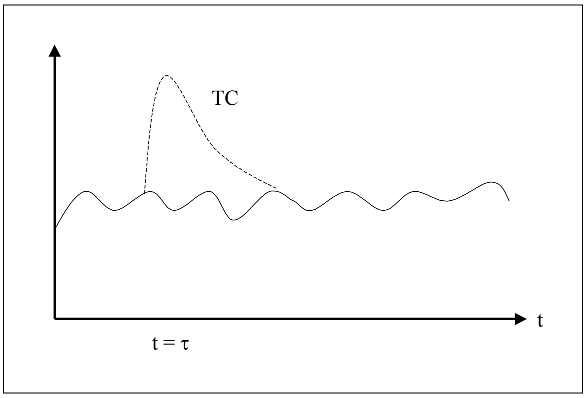
- Level Shifts (LS) that increase or decrease all the observations from a certain time point onward by some constant amount. In this case, the polynomial: $v_i(B) = 1/(1 - B)$.

Fig. 3: Level Shifts



- Temporary Changes (TC) that allow an abrupt increase or decrease in the level of a series which then returns to its previous level exponentially rapidly. Their speeds of decay depend on the parameter $v_i(B) = 1/(1 - \delta B)$, where 0<δ<1.

Fig. 4: Temporary Changes



The literature considers that AOs and IOs are outliers which are related to an exogenous and endogenous change in the series, respectively, and that TCs and LSs are more in the nature of structural changes. TCs represent ephemeral shifts in a series whereas LSs are more the reflection of permanent shocks. However, IOs will have a relatively persistent effect on the level of the series.

An ARIMA model is fitted to y_t^* in (1) and the residuals are obtained:

$$\hat{a}_t = \pi(B)Y_t, (3)$$

where $\pi(B) = \frac{\alpha(B)\phi(B)}{\theta(B)} = 1 - \pi_1 B - \pi_2 B^2 - ...$

For the three types of outliers in (2), the equation in (3) becomes:

AO: $\hat{a}_t = a_t + \omega_1 \pi(B) I_t(\tau)$

IO: $\hat{a}_t = a_t + \omega_2 I_t(\tau)$

LS: $\hat{a}_t = a_t + \omega_3 \left[\frac{\pi(B)}{(1-B)}\right] I_t(\tau)$

TC: $\hat{a}_t = a_t + \omega_2 \left[\frac{\pi(B)}{(1-\delta B)}\right] I_t(\tau)$

These expressions can then be viewed as a regression model for $\hat{a}_t$, i.e.,

$$\hat{a}_t = \omega_i x_{i,t} + a_t$$

with:

for all i and $t < \tau$: $\quad x_{i,t} = 0$

for all i and $t = \tau$: $\quad x_{i,t} = 1$

$$x_{1,t+k} = -\pi_k \qquad \text{(AO);}$$

$$x_{2,t+k} = 0 \qquad \text{(IO);}$$

for $t > \tau$ and $k \geq 1$:
$$x_{3,t+k} = 1 - \sum_{j=1}^{k} \pi_j \qquad \text{(LS);}$$

$$x_{4,t+k} = \delta^k - \sum_{j=1}^{k-1} \delta^{k-j} \pi_j - \pi_k \qquad \text{(TC).}$$

The test statistics for the types of outliers are given by:

$$\text{AO:} \quad \hat{\tau}_1(\tau) = \left[\hat{\omega}_1(\tau)/\hat{\sigma}_a\right] \Big/ \left(\sum_{t=\tau}^{n} x_{1,y}^2 \right)^{1/2}$$

$$\text{IO:} \quad \hat{\tau}_2(\tau) = \hat{\omega}_2(\tau)/\hat{\sigma}_a$$

$$\text{LS:} \quad \hat{\tau}_3(\tau) = \left[\hat{\omega}_3(\tau)/\hat{\sigma}_a\right] \Big/ \left(\sum_{t=\tau}^{n} x_{3,t}^2 \right)^{1/2}$$

$$\text{TC:} \quad \hat{\tau}_4(\tau) = \left[\hat{\omega}_4(\tau)/\hat{\sigma}_a\right] \Big/ \left(\sum_{t=\tau}^{n} x_{4,t}^2 \right)^{1/2}$$

$$\hat{\omega}_i(\tau) = \sum_{t=\tau}^{n} \hat{a}_t x_{i,t} \Big/ \sum_{t=\tau}^{n} x_{i,y}^2$$

for $i = 1,3,4$ and $\hat{\omega}_2(\tau) = \hat{a}_t$

where $\hat{\omega}_i(\tau)(i = 1-4)$ denotes the estimation of the outlier impact at time $t = \tau$, and $\hat{\sigma}_a$ is an estimate of the variance of the residual process.

An outlier is identified at time $t = \tau$ when the test statistics $\hat{\tau}_i(\tau)$ exceeds a critical value. In TRAMO (Time Series Regression with ARIMA Noise, Missing Observations, and Outliers) the critical value is determined by the number of observations in the series based on simulation experiments. The different test statistics at time $t = \tau$ are compared in order to identify the type of outlier. The one chosen has the greatest significance such as $\hat{\tau}_{max} = \max\left|\hat{\tau}_i(\tau)\right|$.

When an outlier is detected, we can adjust the observation Y_t at time $t = \tau$ to obtain the corrected Y_t^* via (2) using the $\hat{\omega}_i$, i.e. $Y_t^* = Y_t - \hat{\omega}_i v_i I_t(\tau)$. Finally, the procedure is repeated until no outlier is detected. A multiple regression on Y_t^* is performed on the various outliers detected to identify spurious outliers.

References

Darné, O. & Diebolt, C. [2004], "Unit Roots and Infrequent Large Shocks: New International Evidence on Output", *Journal of Monetary Economics*, 51 (7), pp. 1449-1465.

Darné, O. & Diebolt, C. [2005], "Non-Stationarity Tests in Macroeconomic Time Series", in: Diebolt, C. & Kyrtsou, C. in collaboration with Darné, O. (Eds.): *New Trends in Macroeconomics*, Springer, Berlin, pp. 173-194.

Darné, O. & Diebolt, C. [2006], "Chocs temporaires et permanents dans le PIB de la France, du Royaume-Uni et des Etats-Unis", *Revue d'Economie Politique*, 116 (1), pp. 65-78.

Metz, R. [2002], *Trend, Zyklus und Zufall. Bestimmungsgründe und Verlaufsformen langfristiger Wachstumsschwankungen*, Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte, Beihefte Nr. 165, Franz Steiner Verlag, Stuttgart.

Metz, R. [2004], "Stochastic Trends in German Gross Domestic Product (GDP) 1850-1990", in: Metz, R., Lösch, M. & Edel, K. (Eds.): *Zeitreihenanalyse in der Empirischen Wirtschaftsforschung*, Lucius & Lucius, Stuttgart, pp. 147-174.