Advances in historical time series analysis
Diebolt, Claude

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Nutzungsbedingungen:
Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
https://creativecommons.org/licenses/by/4.0/deed.de

Terms of use:
This document is made available under a CC BY Licence (Attribution). For more Information see:
https://creativecommons.org/licenses/by/4.0

Diese Version ist zitierbar unter / This version is citable under:
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-62818
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

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:

$$y_t = y_t^* + \sum_i \omega_i \nu_i(B)I_t(\tau) \quad (2)$$

where $y_t^*$ is an ARIMA process, $\nu_i(B)$ is the polynomial characterizing the outlier occurring at time $t = \tau$, $\omega_i$ represents its impact on the series and $I_t(\tau)$ is an indicator function with the value 1 at time $t = \tau$ 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: $\nu_i(B) = 1$.

  ![Additive Outliers](image)

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

320
- 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: $v_t(B) = 1/(1 - B)$.

- 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 $v_t(B) = 1/(1 - \delta B)$, where $0 < \delta < 1$. 
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:

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

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: \quad \hat{a}_t = a_t + \omega_0 \pi(B)I_t(\tau)

IO: \quad \hat{a}_t = a_t + \omega_0 I_t(\tau)

LS: \quad \hat{a}_t = a_t + \omega_0 \left[ \frac{\pi(B)}{1 - B} \right] I_t(\tau)

TC: \quad \hat{a}_t = a_t + \omega_0 \left[ \frac{\pi(B)}{1 - B} \right] I_t(\tau)

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

$$\hat{a}_t = \omega_0 x_{t,\tau} + a_t$$

with:
for all $i$ and $t < \tau$: 
$$x_{i,t} = 0$$
for all $i$ and $t = \tau$: 
$$x_{i,t} = 1$$

$$x_{i,t+k} = -\pi_k$$

$$x_{2,t+k} = 0$$ (AO);

for $t > \tau$ and $k \geq 1$:
$$x_{3,t+k} = 1 - \sum_{j=1}^{k} \pi_j$$

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

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

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

IO: 
$$\hat{r}_2(\tau) = \frac{\hat{\omega}_2(\tau)}{\hat{\sigma}_a}$$

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

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

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

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

where $\hat{\omega}_i(\tau)(i = 1 \ldots 4)$ 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 $\hat{r}_j(\tau)$ 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 $\hat{r}_{\text{max}} = \max_{j=1}^{4} \hat{r}_j(\tau)$.

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. $Y^*_t = Y_t - \hat{\omega}_i Y_{\tau}$. 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


