Time series clustering based on nonparametric forecast

Vilar, J.A.[1], Alonso, A.M.[2] and Vilar, J.M.[1]

[1]Departamento de Matemáticas
Universidade da Coruña, Spain

[2]Departamento de Estadística
Universidad Carlos III de Madrid, Spain

URJC, Madrid, Spain

Outline

The problem

Time series clustering problems arise when we observe a sample of time series and we want to group them into different categories or clusters.

This a central problem in many application fields and hence time series clustering is nowadays an active research area in different disciplines including finance and economics, medicine, engineering, seismology and meteorology, among others.

Dissimilarity measures

Key point

The metric chosen to assess the dissimilarity between two data objects plays a crucial role in time series clustering.

Dissimilarity measures

Different dissimilarity criteria specifically designed to deal with time series have been proposed in the literature. Some examples are:

Piccolo, D. A distance measure for classifying ARIMA models. J. Time Ser. Anal., 11(2), 153-164 (1990)

Kakizawa, Y., Shumway, R.H. and Taniguchi, M. Discrimination and clustering for multivariate time series. J. Am. Stat. Assoc., **93**, 328-340, (1998)

Maharaj, E.A. A significance test for classifying ARMA models. J. Statist. Comput. Simul., **54**, 305-331 (1996)

Vilar, J.A. and Pértega, S. Discriminant and cluster analysis for Gaussian stationary processes: Local linear fitting approach. J. Nonparametr. Stat., **16**(3-4), 443-462 (2004)

Caiado, J., Crato, N. and Peña, D. A periodogram-based metric for time series classification. Comput. Statist. Data Anal. 50, 2668-2684 (2006)

Corduas, M., Piccolo, D. Time series clustering and classification by the autoregressive metric. Comput. Statist. Data Anal. **52**, 1860–1872 (2008)

Dissimilarity measures

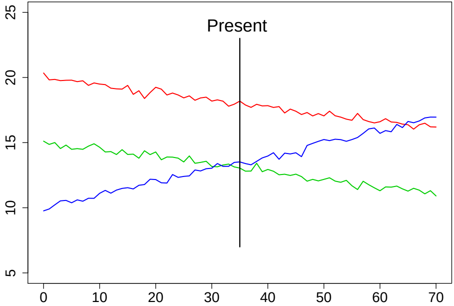
Conceptually most of the dissimilarity criteria proposed for time series clustering lead to a notion of similarity relying on two possible criteria:

- Proximity between raw series data
- Proximity between underlying generating processes

In both cases, the classification task becomes inherently static since similarity searching is governed only by the behavior of the series over their periods of observation.

A dissimilarity measure based on the forecast densities

In many practical situations, the real interest of clustering is the long term behavior and, in particular, on how the forecasts at a specific horizon can be grouped.



A dissimilarity measure based on the forecast densities

Alonso *et al.* (2006) propose a dissimilarity measure based on comparing the full forecast densities associated to each series in the sample.

Remarks:

- In practice, the forecast densities are unknown.
- They suggest to approximate them by using a smoothed sieve bootstrap procedure combined with kernel density estimation techniques.
- Such a procedure requires the assumption that the considered time series admit an $AR(\infty)$ representation.

Our approach

Objective

Extend the clustering procedure proposed by Alonso et al. (2006) to cover the case of nonparametric models of arbitrary autoregressions.

Our approach

1. Our approach does not assume any parametric model for the true autoregressive structure of the series.
2. The mechanism to obtain bootstrap predictions is based on mimicking the generating process using a nonparametric estimator of the autoregressive function and a bootstrap resample of the nonparametric residuals.
3. In this way, we provide an useful device for classifying nonlinear autoregressive time series, including extensively studied parametric models (*TAR* models, *EXPAR* models, *STAR* models, bilinear models, . . .).

The clustering procedure: General description

A general class of autoregressive processes

Let Ξ be the class of real valued stationary processes $\{X_t\}_{t\in\mathbb{Z}}$ such that

$$X_t = m(\boldsymbol{X}_{t-1}) + \varepsilon_t,$$

where

- $\{\varepsilon_t\}$ is an i.i.d. sequence
- $\boldsymbol{X}_{t-1}$ is a d-dimensional vector of known lagged variables
- $m(\cdot)$ is assumed to be a smooth function but it is not restricted to any pre-specified parametric model.

The clustering procedure: General description

Specific clustering problem

We wish to perform a cluster analysis on a set S of s time series belonging to Ξ.

In other terms, $S = \left\{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(s)}\right\}$, where, for $i = 1, \ldots, s$, $\mathbf{X}^{(i)} = \left(X_1^{(i)}, \ldots, X_T^{(i)}\right)$ is generated from a process satisfying the above mentioned model.

Clustering principle

The cluster solution must capture similarities in the behaviors of the predictions at a prefixed horizon: $T + b$.

The clustering procedure: General description

Dissimilarity measures

The dissimilarity between $\mathbf{X}^{(i)}$ and $\mathbf{X}^{(j)}$ is measured by:

$$
\begin{aligned}
D_{1,ij} &= \int \left| f_{X^{(i)}_{T+b}}(x) - f_{X^{(j)}_{T+b}}(x) \right| dx \\
D_{2,ij} &= \int \left(f_{X^{(i)}_{T+b}}(x) - f_{X^{(j)}_{T+b}}(x) \right)^2 dx
\end{aligned}
$$

where $f_{X^{(i)}_{T+b}}(\cdot)$ denotes the density of the forecast $X^{(i)}_{T+b}$.

The clustering procedure: General description

Remark

$D_{2,ij}$ presents a serious drawback to perform cluster analysis

If $\{x : f_{X^{(i)}_{T+b}}(x) > \varepsilon\} \cap \{x : f_{X^{(j)}_{T+b}}(x) > \varepsilon\} = \emptyset$, for a sufficiently small $\varepsilon > 0$, then:

$$D_{2,ij} \approx \int f^2_{X^{(i)}_{T+b}}(x)dx + \int f^2_{X^{(j)}_{T+b}}(x)dx.$$

The clustering procedure: General description

Computation of bootstrap approximations to the distances $D_{u,ij}$

In practice, distances $D_{u,ij}$ are consistently approximated by replacing the unknown $f_{X_{T+b}^{(i)}}$ by kernel-type density estimates $\hat{f}_{X_{T+b}^{(i)}}$ constructed on the basis of bootstrap predictions, that is

$$\hat{D}_{u,ij}^{*} = \int \left| \hat{f}_{X_{T+b}^{(i)*}}(x) - \hat{f}_{X_{T+b}^{(j)*}}(x) \right|^{u} dx, \quad i,j = 1, \ldots, s,$$

for $u = 1, 2$.

The clustering procedure: General description

Application of a agglomerative hierarchical cluster algorithm

Once the pairwise dissimilarity matrix $\hat{D}_u^* = \left(\hat{D}_{u,ij}^*\right)$ is obtained, a standard agglomerative hierarchical clustering algorithm based on $\hat{D}_u^*$ is carried out.

Generating bootstrap predictions

Bootstrap algorithm (*Autoregression Bootstrap*)

1. Estimate m using a Nadaraya-Watson estimator $\hat{m}_{g_1}$.
2. Compute the nonparametric residuals, $\widehat{\varepsilon}_t = X_t - \hat{m}_{g_1}(\boldsymbol{X}_{t-1})$.
3. Construct a kernel estimate, $\hat{f}_{\tilde{\varepsilon},h}$, of the density function associated to the centered residuals $\tilde{\varepsilon}_t = \widehat{\varepsilon}_t - \widehat{\varepsilon}_\bullet$.
4. Draw a bootstrap-resample ε_t^* of i.i.d. data from $\hat{f}_{\tilde{\varepsilon},h}$.
5. Define the bootstrap series X_t^*, by the recursion $X_t^* = \hat{m}_{g_1}(\boldsymbol{X}_{t-1}^*) + \varepsilon_t^*$
6. Obtain the bootstrap autoregressive function, $\hat{m}_{g_2}^*$, using the bootstrap sample $(X_1^*, \ldots, X_T^*)$.
7. Compute bootstrap prediction-paths by the recursion

$$X_t^* = \hat{m}_{g_2}^*(\boldsymbol{X}_{t-1}^*) + \varepsilon_t^*,$$

for $t = T+1, \ldots, T+b$, and $X_t^* = X_t$, for $t \leq T$.

Generating bootstrap predictions

If Steps (5) and (6) are omitted and the prediction-paths in Step (7) are computed using $\hat{m}_{g_1}$ instead of $\hat{m}^{\star}_{g_2}$, then our resampling plan is a conditional bootstrap procedure.

Bootstrap algorithm (*Conditional Bootstrap*)

1. Estimate m using a Nadaraya-Watson estimator $\hat{m}_{g_1}$.
2. Compute the nonparametric residuals, $\widehat{\varepsilon}_t = X_t - \hat{m}_{g_1}(\boldsymbol{X}_{t-1})$.
3. Construct a kernel estimate, $\hat{f}_{\tilde{\varepsilon},h}$, of the density function associated to the centered residuals $\tilde{\varepsilon}_t = \widehat{\varepsilon}_t - \widehat{\varepsilon}_{\bullet}$.
4. Draw a bootstrap-resample ε^*_t of i.i.d. data from $\hat{f}_{\tilde{\varepsilon},h}$.
5. Compute bootstrap prediction-paths by the recursion

$$X^*_t = \hat{m}_{g_1}(\boldsymbol{X}_{t-1}) + \varepsilon^*_t,$$

for $t = T+1, \ldots, T+b$.

Simulation study: Motivation

- Our clustering procedure is aimed at grouping time series with similar forecast densities at a future time $T+b$.

- Under this similarity principle, the efficiency of our clustering procedure relies on the closeness to zero of the values:

$$d_{u,\mathbf{X}^{(\mathbf{i})}} = \int \left| \hat{f}_{X_{T+b}^{(i)*}}(x) - f_{X_{T+b}^{(i)}}(x) \right|^u dx, \quad i = 1, \ldots, s.$$

- Note that if $d_{u,\mathbf{X}^{(i)}} \approx 0$, for all i, then $D^*_{u,ij} \approx D_{u,ij}$, for all i,j, and hence the cluster solutions obtained from both dissimilarity matrices should be close as well.

Simulation study: Simulation features

Autoregressive models (see Luukonen *et al.* (1988) and Giordano *et al.*(2007))

M1	AR	$X_t = 0{,}6X_{t-1} + \varepsilon_t$
M2	Bilinear	$X_t = (0{,}3 - 0{,}2\varepsilon_{t-1})\,X_{t-1} + 1{,}0 + \varepsilon_t$
M3	EXPAR	$X_t = \left(0{,}9\exp\left(-X_{t-1}^2\right) - 0{,}6\right)X_{t-1} + 0{,}3 + \varepsilon_t$
M4	EXPAR	$X_t = \left(0{,}9\exp\left(-X_{t-1}^2\right) - 0{,}6\right)X_{t-1} + 1{,}0\varepsilon_t$
M5	SETAR	$X_t = (0{,}3X_{t-1} - 1{,}0)\,I\,(X_{t-1} \geq 0{,}2) - (0{,}3X_{t-1} + 0{,}5)\,I\,(X_{t-1} < 0{,}2) + \varepsilon_t$
M6	SETAR	$X_t = (0{,}3X_{t-1} + 1{,}0)\,I\,(X_{t-1} \geq 0{,}2) - (0{,}3X_{t-1} - 1{,}0)\,I\,(X_{t-1} < 0{,}2) + \varepsilon_t$
M7	NLAR	$X_t = 0{,}7\,\lvert X_{t-1}\rvert\,(2 + \lvert X_{t-1}\rvert)^{-1} + \varepsilon_t$
M8	STAR	$X_t = 0{,}8X_{t-1} - 0{,}8X_{t-1}\,(1 + \exp(-10X_{t-1}))^{-1} + \varepsilon_t$

In all cases, ε_t consisted of i.i.d. zero-mean variables with variance σ^2.

Simulation study: Simulation features

Distributions for innovations ε_t

(i) Gaussian innovations with unit variance
(ii) Student-t innovations with 3 degrees of freedom
(iii) Centered exponential $Exp(1) - 1$ innovations

Forecast horizons $T + b$

The forecast densities are estimated at a specific forecast horizon $T + b$, with b denoting the number of steps ahead.

Here, we present the results for:

$b = 1$ (short term) $b = 3$ (intermediate term) $b = 10$ (long term)

Simulation study: Simulation features

Resampling methods to generate bootstrap predictions

1. *Autoregression bootstrap* (AB).

$$X_t^* = \hat{m}_g^\star(\boldsymbol{X}_{t-1}^*) + \varepsilon_t^*,$$

where ε_t^* are resampled from a nonparametric approximation of m.

2. *Conditional bootstrap* (CB).

$$X_t^* = \hat{m}_g(\boldsymbol{X}_{t-1}) + \varepsilon_t^*,$$

where ε_t^* are resampled from a nonparametric approximation of m.

3. *AR-sieve bootstrap* (SB) (considered by Alonso *et al.*, 2006).

$X_t^* = \sum_{j=1}^{p} \hat{\phi}_j^\star B^j X_t^\star + \varepsilon_t^*$, where ε_t^* are resampled from a linear approximation of m.

Simulation study: Simulation features

The simulation mechanism (I)

For each of the considered models:

- One thousand time series of length $T = 200$ were simulated.
- With every simulated series $\mathbf{X}$, the b-step-ahead forecast density $f_{X_{T+b}}$ was approximated using:
  - The autoregression bootstrap predictions: $\hat{f}^{AB}_{X^*_{T+b}}$.
  - The conditional bootstrap predictions: $\hat{f}^{CB}_{X^*_{T+b}}$.
  - The *AR*-sieve bootstrap predictions: $\hat{f}^{SB}_{X^*_{T+b}}$.
  - Monte Carlo predictions instead of bootstrap predictions: $\hat{f}^{MC}_{X^*_{T+b}}$.
- Each bootstrap density was estimated using $B = 1000$ replicates.

Simulation study: Simulation features

The simulation mechanism (II)

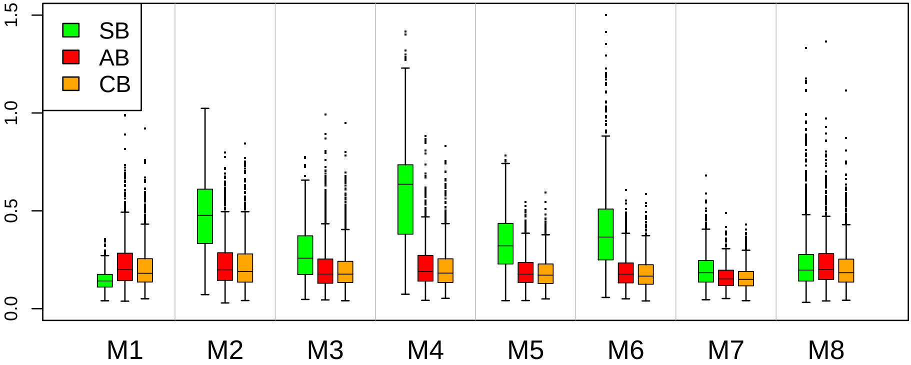
- The behavior with finite samples of the values $d_{u,\mathbf{x}}$ was studied by analyzing the feasible values of

$$d_{u,\mathbf{x}}^{\bullet} = \int \left| \hat{f}^{\bullet}_{X^*_{T+b}}(x) - \hat{f}^{MC}_{X^*_{T+b}}(x) \right|^u dx,$$

for $u = 1, 2$ and $\bullet$ being AB, CB or SB.

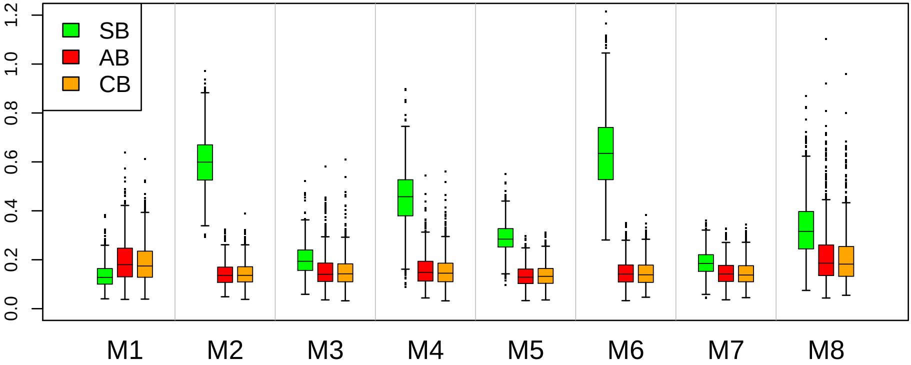
Simulation study: Results

Boxplots of d_1^{SB}, d_1^{AB} and d_1^{CB} for Gaussian innovations and $b = 1$



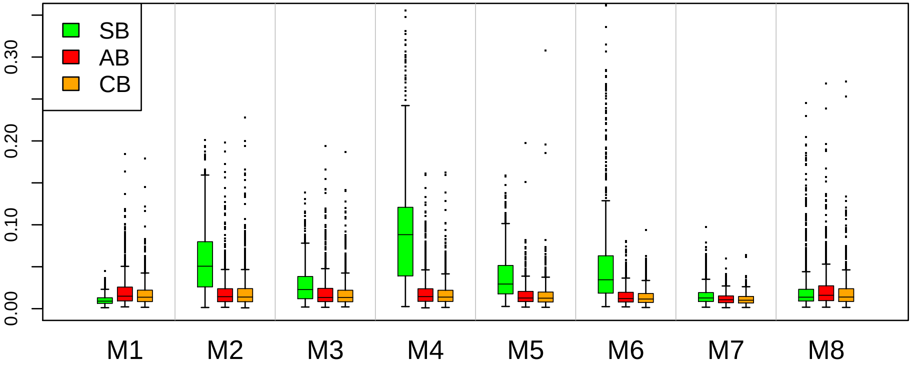
Simulation study: Results

Boxplots of d_1^{SB}, d_1^{AB} and d_1^{CB} for Gaussian innovations and $b = 3$



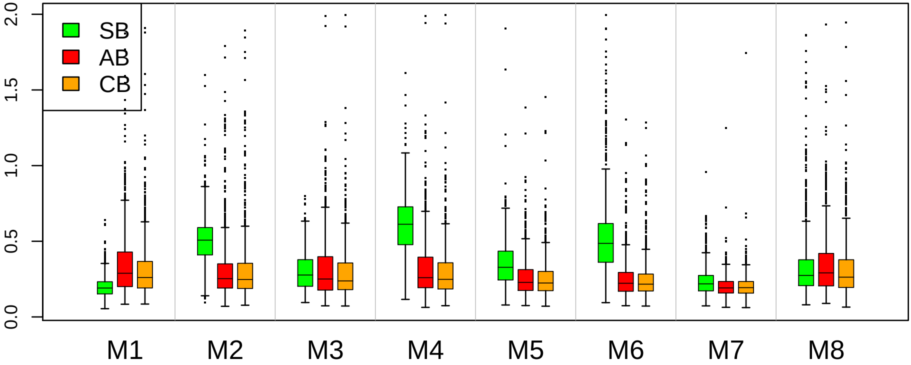
Simulation study: Results

Boxplots of d_2^{SB}, d_2^{AB} and d_2^{CB} for Gaussian innovations and $b = 1$



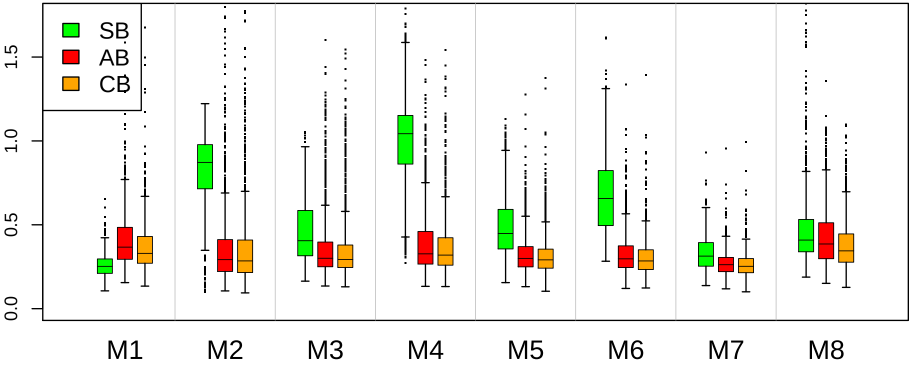
Simulation study: Results

Boxplots of d_1^{SB}, d_1^{AB} and d_1^{CB} for exponential innovations and $b = 1$



Simulation study: Results

Boxplots of d_1^{SB}, d_1^{AB} and d_1^{CB} for t-Student innovations and $b = 1$



Simulation study: Conclusions

Conclusions

1. Under *moderate* or *strong* nonlinearity, the autoregression bootstrap (AB) and the conditional bootstrap (CB) clearly outperform the sieve bootstrap (SB).
2. AB and CB lead to rather similar results although, for some models, a slight improvement is observed when CB is used.
3. Both AB and CB present an interesting *robustness property* with respect to the data generating model.
4. Nonparametric AB and CB procedures are not affected when the nonlinearity is mainly revealed at lags of order higher than one.
5. Previous considerations are valid for both L^1 and L^2 distances, for non Gaussian innovations and for all the considered prediction horizons.

A case-study with real data: Data description

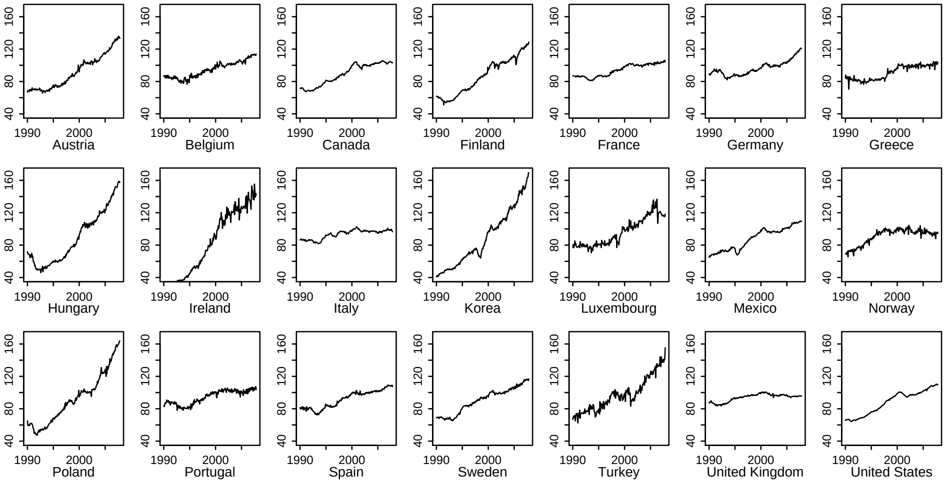
Dataset

A collection of time series representing the monthly industrial production indices (seasonally adjusted) for 21 OECD countries from January 1990 to November 2007. This dataset is available from the Statistics Portal of OECD (http://stats.oecd.org/).

Purpose

To classify the 21 countries in accordance with the performance of their industrial production indices at the next month ($b = 1$).

A case-study with real data: Data description



A case-study with real data: Data description

The series are clearly nonstationary

1. We transformed the series using logarithms (if required) and taking regular differences.
2. The bootstrap prediction-paths for the transformed series were constructed using the CB algorithm.
3. The prediction-paths were then backtransformed to obtain the bootstrap predictions for the original series.

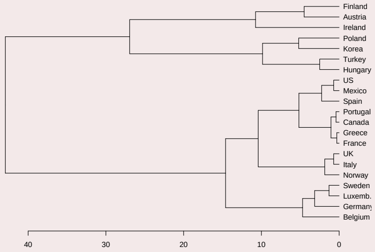
Some of the series are nonlinear

Linearity was rejected at level 0.05 (*McLeod-Li test*) for the series:

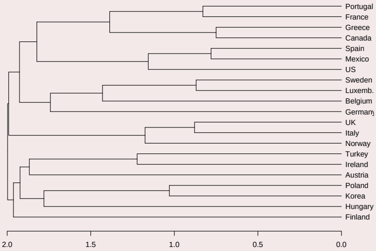
Austria	***Mexico***
Belgium	***Poland***
Canada	***Turkey***
Hungary	***United Kingdom***
Luxembourg	

A case-study with real data: Clustering results

CLUSTERING BASED ON THE LAST OBSERVATION

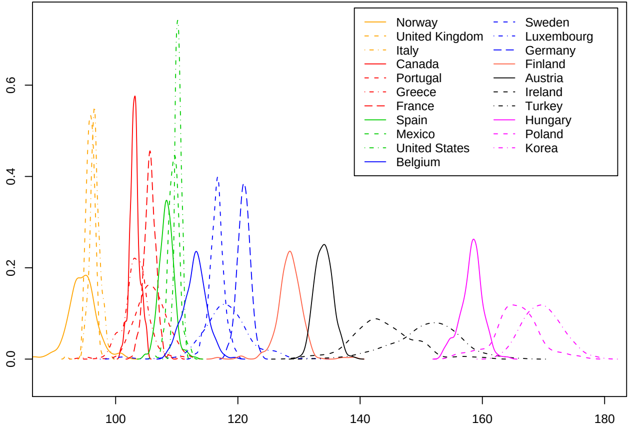


OUR APPROACH: L^1 *distance, cond. bootstrap,* $b = 1$

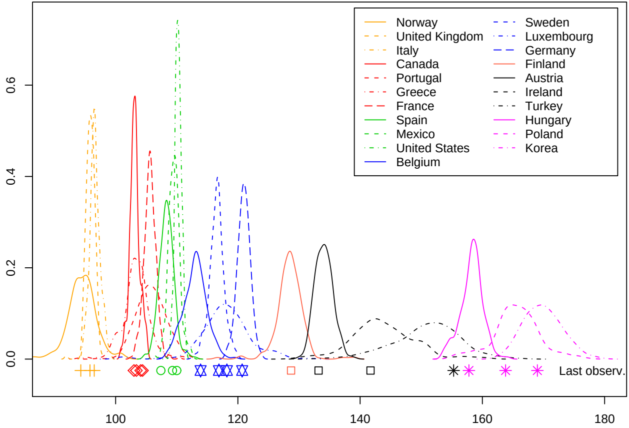


- Similar groups are formed with both procedures.
- Our procedure modified the order in which some series are joined.
- Also, series with the highest indices were regrouped in different way.

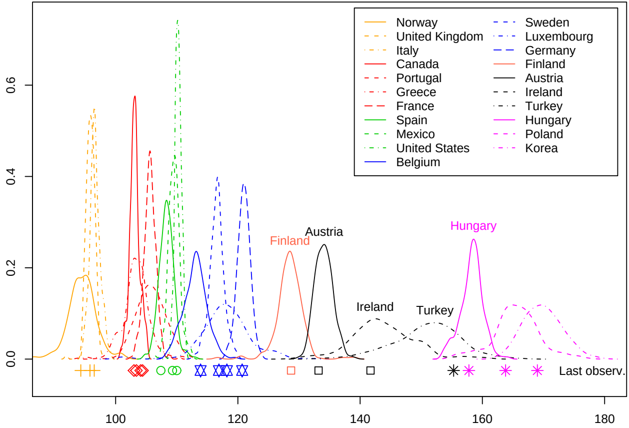
A case-study with real data: Clustering results



A case-study with real data: Clustering results



A case-study with real data: Clustering results

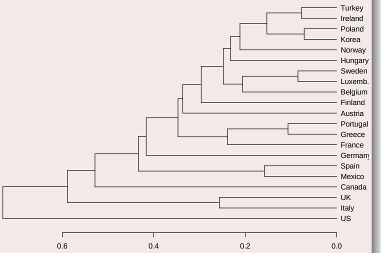


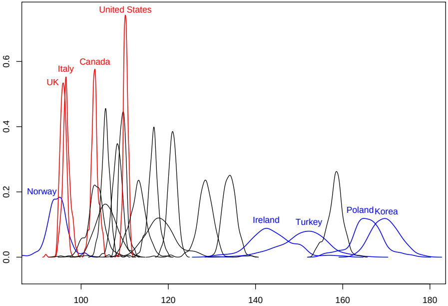
Some illustrative differences:

- **Clustering based on the last observations:**
  - *Turkey* and *Hungary* belong to the same cluster.
  - *Turkey* and *Ireland* are in different clusters.
  - *Finland* and *Austria* are joined very soon.
- **Clustering based on the forecast densities (L^1 dist.):**
  - *Turkey* and *Hungary* are located at different clusters.
  - *Turkey* and *Ireland* are very close and belong to the same cluster.
  - *Finland* and *Austria* remain isolated until very late.

A case-study with real data: Clustering results

OUR APPROACH: L^2 *distance, cond. bootstrap,* $b = 1$





Alonso, A.M., Berrendero, J.R., Hernández, A., Justel, A., 2006. Time series clustering based on forecast densities. Comput. Statist. Data Anal. 51, 762-776.

Corduas, M., Piccolo, D., 2008. Time series clustering and classification by the autoregressive metric. Comput. Statist. Data Anal. 52, 1860-1872.

Chouakria-Douzal, A. & Nagabhushan, P.N., 2007. Adaptive dissimilarity index for measuring time series proximity. Adv. Data Anal. Classif. 1, 5-21.

Franke, J., Kreiss, J.P., Mammen, E., 2002. Bootstrap of kernel smoothing in nonlinear time series. Bernouilli 8, 1-37.

Giordano, F., La Rocca, M., Perna, C., 2007. Forecasting nonlinear time series with neural network sieve bootstrap. Comput. Statist. Data. Anal.

Kakizawa, Y., Shumway, R.H., Taniguchi, M., 1998. Discrimination and clustering for multivariate time series. J. Amer. Statist. Assoc. 93, 328-340.

Luukkonen, R., Saikkonen, P., Terärsvirta, T., 1988. Testing linearity in univariate time series. Scand. J. Statist. 15, 161-175.

Maharaj, E.A., 1996. A significance test for classifying ARMA models. J. Statist. Comput. Simulation 54, 305Ű331.

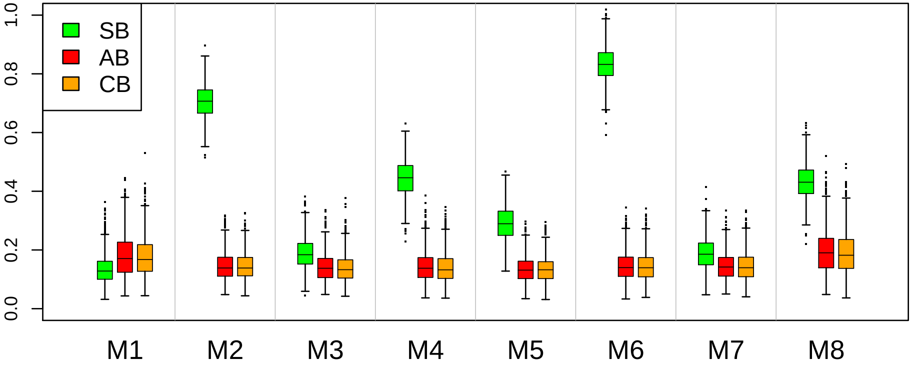
Piccolo, D., 1990. A distance measure for classifying arima models. J. Time Series Anal. 11, 153Ű164.

Tong, H., 1990. Non-linear Time Series: A Dynamical System Approach. Oxford Statistical Science Series, 6.

Vilar, J.A., Pértega, S., 2004. Discriminant and cluster analysis for Gaussian stationary processes: local linear fitting approach. J. Nonparametr. Stat. 16, 443-462.

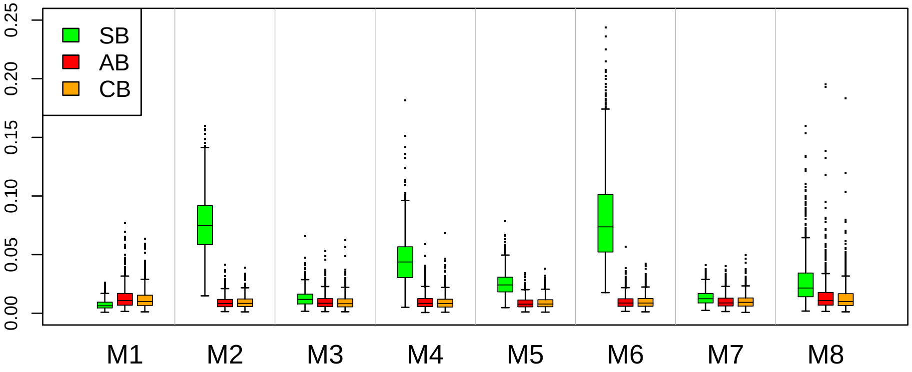
Simulation study: Results

Boxplots of d_1^{SB}, d_1^{AB} and d_1^{CB} for Gaussian innovations and $b = 10$



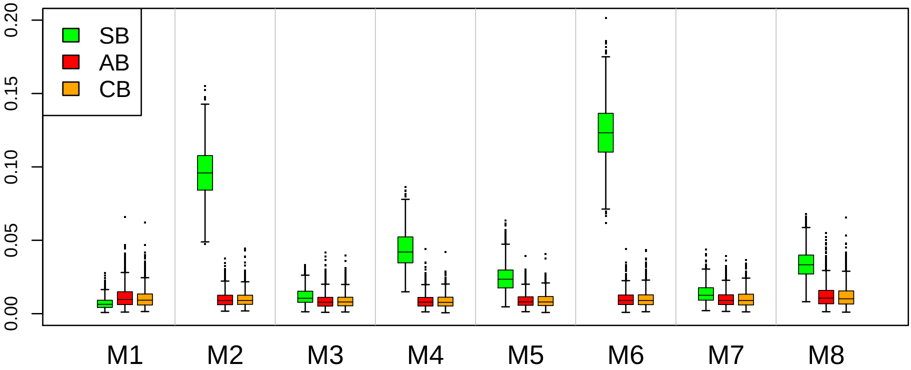
Simulation study: Results

Boxplots of d_2^{SB}, d_2^{AB} and d_2^{CB} for Gaussian innovations and $b = 3$



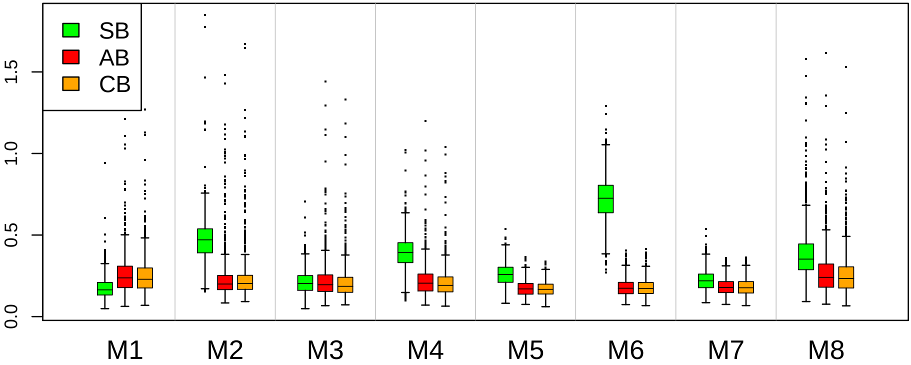
Simulation study: Results

Boxplots of d_2^{SB}, d_2^{AB} and d_2^{CB} for Gaussian innovations and $b = 10$



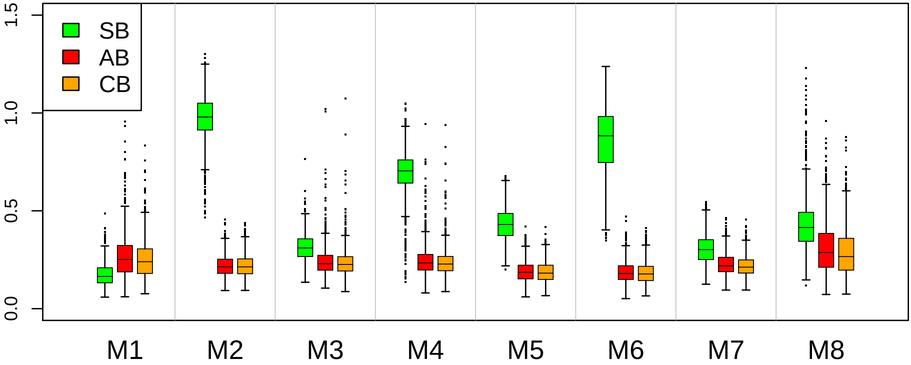
Simulation study: Results

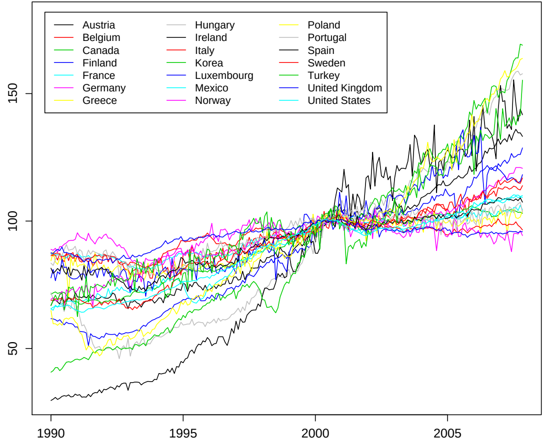
Boxplots of d_1^{SB}, d_1^{AB} and d_1^{CB} for exponential innovations and $b = 3$



Simulation study: Results

Boxplots of d_1^{SB}, d_1^{AB} and d_1^{CB} for t-Student innovations and $b = 3$




Austria
Belgium
Canada
Finland
France
Germany
Greece
Hungary
Ireland
Italy
Korea
Luxembourg
Mexico
Norway
Poland
Portugal
Spain
Sweden
Turkey
United Kingdom
United States
50
100
150
1990
1995
2000
2005