A mixture model for detecting unusual temporal patterns with an application to income modelling

Guangquan Li[1], Nicky Best[1], Philip Clark[2] and Sylvia Richardson[1]

1. Department of Epidemiology and Public Health, Imperial College London, UK
2. Office for National Statistics, UK.

July 1, 2009 SAE2009

Outline

Trends classification

- Temporal trends of average household income are likely to be similar in shape across different administrative regions (e.g., counties and municipalities).
- However, some regions may display unexpected changes over time due to, for example, different responses to policies and/or social changes.
- Detection of areas with "unusual" time trends provides a tool to
  - evaluate impacts of policies;
  - assist allocating resources efficiently to areas that exhibit "downward" trends.

Trends classification

- Temporal trends of average household income are likely to be similar in shape across different administrative regions (e.g., counties and municipalities).
- However, some regions may display unexpected changes over time due to, for example, different responses to policies and/or social changes.
- Detection of areas with "unusual" time trends provides a tool to
  - evaluate impacts of policies;
  - assist allocating resources efficiently to areas that exhibit "downward" trends.

This could be easily done if data on income were available for all households in the country. But ...

Survey setting

- Collecting data on the entire population is costly and time consuming.
- Typical survey constructs samples only on some percentage (e.g., 1%) of the population
- The global/mean temporal trend can be estimated accurately
- But local trend estimations may not be reliable due to variability created by small samples.

A mixture model

- To tackle this problem, we have proposed a Bayesian mixture modelling framework which allows
  1. to extract the common temporal trend;
  2. to provide reliable estimates of the local trends;

A mixture model

- To tackle this problem, we have proposed a Bayesian mixture modelling framework which allows
  1. to extract the common temporal trend;
  2. to provide reliable estimates of the local trends;
  3. hence facilitate detection of “unusual” patterns.

A mixture model

- To tackle this problem, we have proposed a Bayesian mixture modelling framework which allows
  1. to extract the common temporal trend;
  2. to provide reliable estimates of the local trends;
  3. hence facilitate detection of "unusual" patterns.
- We will use a Swedish income data to evaluate the performance of our mixture model for trend classification.

Data Description

- The data set is from the LOUISE Population registry in Sweden from 1992 to 1999.
- It contains various socio-economic variables at both the individual and household levels.
- The variable of interest is the average equivalised income per household (AvgEqInc/HH) for each of the 284 municipalities in the country.

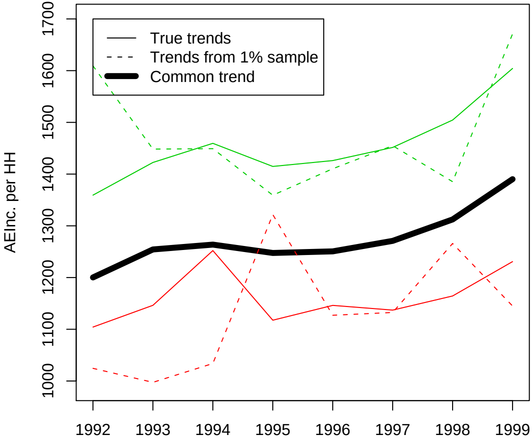
Data Description

- The data set is from the LOUISE Population registry in Sweden from 1992 to 1999.
- It contains various socio-economic variables at both the individual and household levels.
- The variable of interest is the average equivalised income per household (AvgEqInc/HH) for each of the 284 municipalities in the country.
- We will NOT use the entire data set.

Data Description

- The data set is from the LOUISE Population registry in Sweden from 1992 to 1999.
- It contains various socio-economic variables at both the individual and household levels.
- The variable of interest is the average equivalised income per household (AvgEqInc/HH) for each of the 284 municipalities in the country.
- We will NOT use the entire data set.
- Instead we use data from a mock survey in which we sample data on income for only **1%** of the total number of households in each municipality in each year.
- The number of HHs in the 1% data ranges from 14 to 307.

Examples of trends: "True" vs. sample



Model specification

- We employ the Bayesian hierarchical modelling approach.
- Denoting $\hat{\hat{Y}}_{i,t}$ and $\hat{\sigma}^2_{i,t}$ the AvgEqInc/HH and the corresponding design variance of municipality i in year t obtained from a typical 1% sample.
  - First level
$$\hat{\hat{Y}}_{i,t} \sim N(\mu_{i,t}, \hat{\sigma}^2_{i,t}) \tag{1}$$
  - Second level
$$\mu_{i,t} = \alpha + \beta \cdot \bar{X}_{i,t} + \mathrm{mix}_{i,t} \tag{2}$$

Spec. for the mixture term

- For fitting a local trend, we use either a trend that is common for all areas or a trend that is specific to that area.
- For the common trend model, we use a first order random walk.
- For the area-specific model, we use a piecewise linear spline model (PWLS).

Spec. for the mixture term

- For fitting a local trend, we use either a trend that is common for all areas or a trend that is specific to that area.
- For the common trend model, we use a first order random walk.
- For the area-specific model, we use a piecewise linear spline model (PWLS).
- We introduce a latent allocation variable z_i for the selection:

$$\text{mix}_{i,t}|z_i = \begin{cases} RW_t + \eta_i + \epsilon_{i,t} & \text{if } z_i = 1 \text{ (common)} \\ \mathrm{PWLS}_{i,t} & \text{if } z_i = 0 \text{ (area-spec.)} \end{cases} \tag{3}$$

Spec. for the mixture term

- For fitting a local trend, we use either a trend that is common for all areas or a trend that is specific to that area.
- For the common trend model, we use a first order random walk.
- For the area-specific model, we use a piecewise linear spline model (PWLS).
- We introduce a latent allocation variable z_i for the selection:

$$\text{mix}_{i,t}|z_i = \begin{cases} RW_t + \eta_i + \epsilon_{i,t} & \text{if } z_i = 1 \text{ (common)} \\ \mathrm{PWLS}_{i,t} & \text{if } z_i = 0 \text{ (area-spec.)} \end{cases} \qquad (3)$$

where

- $\text{RW}_t \sim$ Random Walk of order 1 (common);
- $\text{PWLS}_{i,t} \equiv \nu_i + f_i(\kappa_i; t)$ (area-spec. smooth model).
- $z_i \sim Bern(0.5)$ a *priori*

Spec. for the mixture term *cnt.*

$$\mathrm{mix}_{i,t} = z_i \cdot (\eta_i + RW_t + \epsilon_{i,t}) + (1 - z_i) \cdot \mathrm{PWLS}_{i,t}$$

- $\eta_{1:N} \sim$ improper Conditional Autoregressive (spatial) model (*ICAR*) is an additive local adjustment to the overall level of the common trend;

Spec. for the mixture term *cnt.*

$$\text{mix}_{i,t} = z_i \cdot (\eta_i + RW_t + \epsilon_{i,t}) + (1 - z_i) \cdot \text{PWLS}_{i,t}$$

- $\eta_{1:N} \sim$ improper Conditional Autoregressive (spatial) model (*ICAR*) is an additive local adjustment to the overall level of the common trend;
- $\epsilon_{i,t}$ creates a (small) buffer for the common trend, allowing for some flexibility.
- Three specifications are considered for $\epsilon_{i,t}$:
  1. $\epsilon_{i,t} = 0$ (Model 1);
  2. $\epsilon_{i,t} \sim N(0, \sigma^2)$ (Model 2);
  3. $\epsilon_{i,t} \sim N(0, \sigma_t^2)$ (Model 3).

Some remarks

$$\text{mix}_{i,t} = z_i \cdot (\eta_i + RW_t + \epsilon_{i,t}) + (1 - z_i) \cdot (\nu_i + f_i(\kappa_i; t))$$

1. The terms η_i and ν_i should not be combined as they play different roles in the model;
2. A sum-to-zero constraint is imposed on $\text{RW}_{1:T}$; η_i is added to adjust $\text{RW}_{1:T}$ additively to fit data;
3. But ν_i is the local intercept for the spline model.

Some remarks

$$\mathrm{mix}_{i,t} = z_i \cdot (\eta_i + RW_t + \epsilon_{i,t}) + (1 - z_i) \cdot (\nu_i + f_i(\kappa_i; t))$$

1. The terms η_i and ν_i should not be combined as they play different roles in the model;
2. A sum-to-zero constraint is imposed on $\mathrm{RW}_{1:T}$; η_i is added to adjust $\mathrm{RW}_{1:T}$ additively to fit data;
3. But ν_i is the local intercept for the spline model.
4. In fact, an area where both the common and the local models fit equally well, η_i is approximately equal to the mean of the corresponding spline ($\mathbb{E}_t(PWLS_{i,\cdot})$).

A computation issue

- Model fitting is done in WinBUGS.
- In particular fitting of the local splines is done using the reversible jump interface in WinBUGS, which allows us to treat the number of knots, the positions of the knots and other spline coefficients. (This is difficult!)

A computation issue

- Model fitting is done in WinBUGS.
- In particular fitting of the local splines is done using the reversible jump interface in WinBUGS, which allows us to treat the number of knots, the positions of the knots and other spline coefficients. (This is difficult!)
- As a result, standard implementation of mixture models did not work because the model always selected the "easy fitting" RW model.

A computation issue

- Model fitting is done in WinBUGS.
- In particular fitting of the local splines is done using the reversible jump interface in WinBUGS, which allows us to treat the number of knots, the positions of the knots and other spline coefficients. (This is difficult!)
- As a result, standard implementation of mixture models did not work because the model always selected the "easy fitting" RW model.
- Inspired by Carlin and Chib (1995), we implemented an algorithm that runs both trend models in parallel.
- We can then select one or the other at each iteration.
- This implementation speeds up convergence!

Recap

- Purpose: to detect unusual temporal trends of income;
- Data: from a mock survey which covers only 1% of the total number of households in each year;
- Model: the Bayesian mixture model;
- Verification: the "true" unusual trends from the entire data.

Defining the "true" unusual trends from all data

- The entire data set will be used here to define the true unusual trends.
- To assess the similarity of the local and the mean trends in shape, we define

$$R_{i,t} = \frac{\text{Local trend of area } i}{\text{National mean trend}}$$

- To remove the possible difference between the local and mean trends, we standardize the ratios $R_{i,t}$ by its mean, i.e.,$r_{i,t} = |R_{i,t} - \mathbb{E}_t(R_{i,\cdot})|$.
- If the pattern of a local trend is similar to that of the mean trend,

$$r_{i,t} \approx 0 \text{ for all } t = 1, ..., 8.$$

- If the local trend is unusual, then $r_{i,t} >> 0$ for some t.

Defining "true" unusual trends *cnt.*

- For each area, we define a critical value, α_i, which is the average of the N largest $r_{i,t}$ (out of 8).
- Then

$$\begin{cases} \text{the trend of the } i^{th} \text{ area is similar to the common} & \text{if } \alpha_i < \alpha_{cut} \\ \text{the trend of the } i^{th} \text{ area is unusual} & \text{if } \alpha_i > \alpha_{cut} \end{cases}$$

- To define α_{cut}, we take the q^{th} percentile (e.g., the 90^{th} percentile) of $r_{i,t}$.
- The bigger the N is, the more time points that the local trend differs from the common.
- The higher percentile q we take, the larger the differences will be between the local and the common trends.

Summary of the “true” unusual areas

Magnitude of the difference

Number of unusual time points

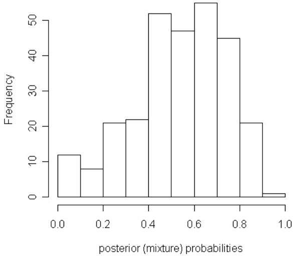
	$q = 90$	$q = 95$	$q = 99$
3 years	66	30	3
4 years	51	22	3
5 years	35	18	2
6 years	24	12	2

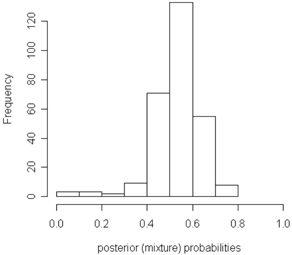
- Purpose: to detect unusual temporal trends of income;
- Data: from a mock survey which covers only 1% of the total number of households in each year;
- Model: the Bayesian mixture model;
- Verification: the "true" unusual trends from the entire data.

- Purpose: to detect unusual temporal trends of income;
- Data: from a mock survey which covers only 1% of the total number of households in each year;
- Model: the Bayesian mixture model;
- Verification: the "true" unusual trends from the entire data.

Results...

Histograms of the posterior mean of $z_i(p_i = \mathbb{E}(z_i))$

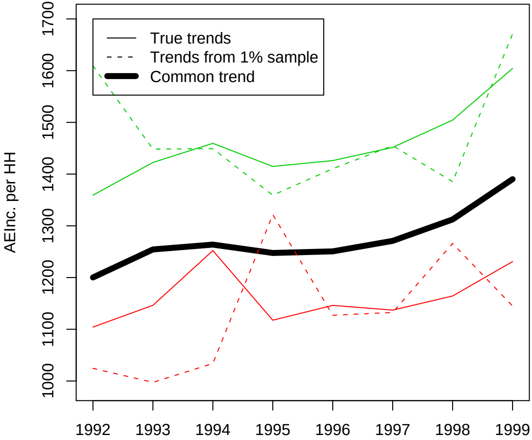




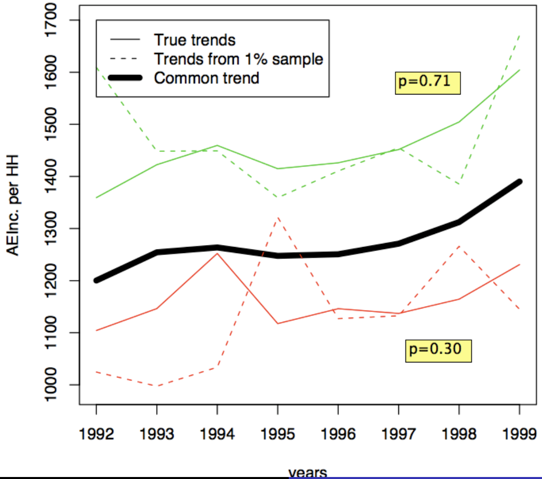
With buffer (V2) | **One replicate**

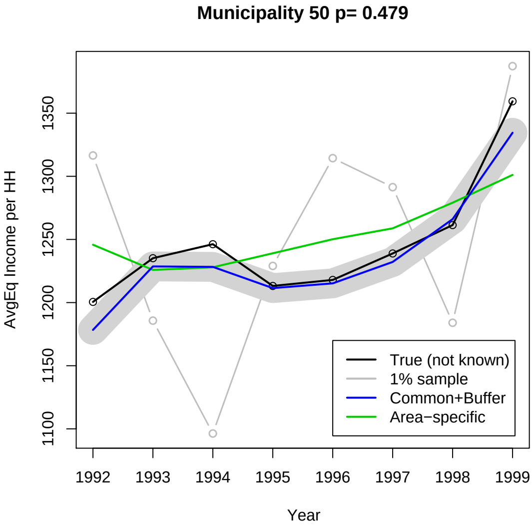
With buffer (V2) | **20 replicates**

Examples of trends: "True" vs. sample



Correct classifications




Municipality 50 p= 0.479
AvgEq Income per HH
1100
1150
1200
1250
1300
1350
1992
1993
1994
1995
1996
1997
1998
1999
Year
True (not known)
1% sample
Common+Buffer
Area−specific

ROC curves (3 Years 90%)

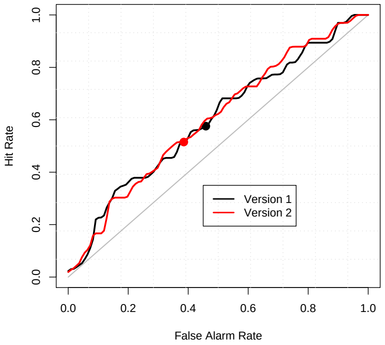


Figure: One typical replicate

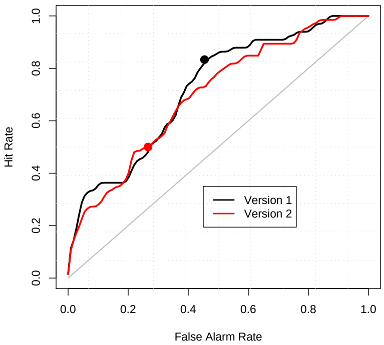


Figure: 20 replicates

ROC curves (3 Years 95%)

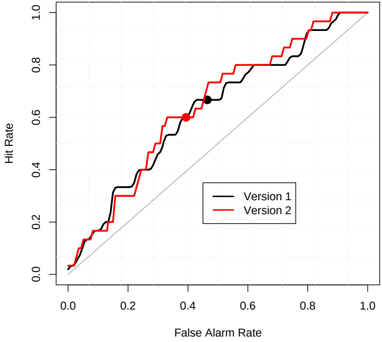


Figure: One typical replicate

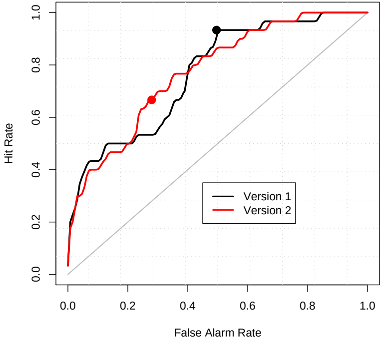


Figure: 20 replicates

ROC curves (5 Years 95%)

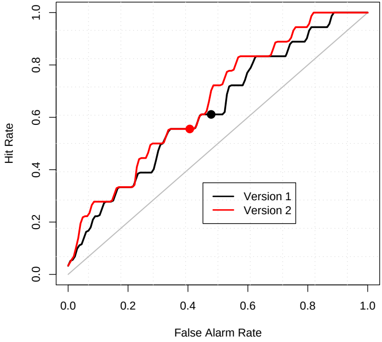


Figure: One typical replicate

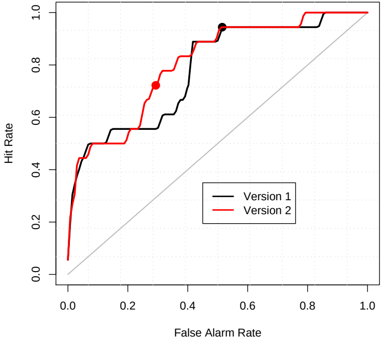


Figure: 20 replicates

Conclusions

- The classification ability of the model depends on
  - the number of time points that the local trend differs from the common one;
  - the magnitude of the difference.
- The model performed well in detecting trends with departure from the common in three or more years and/or the departures are large.
- Although it is rare to have more than one replicate in practice, replicates reduce the uncertainty and hence provide better classification performance from the model.
- This results are anticipated because the model is constructed to examine the difference of the overall pattern rather tahn depicting departures at individual time points.

Discussions

- The current model setting can be easily extended to detect differences at individual time points, for example, replacing z_i by $z_{i,t}$.
- A similar approach was proposed in the context of disease mapping by Abellan et al. (2008) who assumed a separable space-time model and classified areas with unusual trends by assessing the residuals.
- Compare performance of trend classification using different ways of decomposing the space-time variability.
- Extensions to Poisson data, e.g., to model rates of unemployment/crime trends/disease trends etc.

Refereces

- J.J. Abellan, S. Richardson, and N. Best. Use of space-time models to investigate the stability of patterns of disease. Environmental Health Perspectives, 116(8):1111-1119, 2008.
- B.P. Carlin and S. Chib. Bayesian model choice via Markov chain Monte Carlo methods. JRSSB, 57(3): 473-484, 1995.