ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

BACHELOR THESIS

ECONOMETRICS AND OPERATIONS RESEARCH

ACADEMIC YEAR 2017-2018

Heterogeneous Structural Breaks in Panel Data Models: Application on Population Growth

Author
T.N.Q. TRAN
432825tt

Supervisor
Dr. W. WANG
Second Assessor
A. Castelein, MSc

July 7, 2018

Abstract

This paper revisits the hybrid estimation method, GAGFL, which is a combination of the grouped fixed effects approach and the adaptive group fused Lasso, described by Okui & Wang (2018). This estimation method takes into account the heterogeneity of individuals as well as heterogeneity of coefficient estimates in panel data. It aims to model individual heterogeneity by estimating an underlying grouped pattern in the data, such that all the individuals within a group share the same slope coefficient estimates. Allowing for multiple structural breaks in the regression coefficients models the heterogeneity of the slope coefficients. However, the break date estimates as well as the number of breaks can differ among the groups. By means of a Monte Carlo simulation, it is shown that this method performs well in finite samples. Many studies have been conducted relating to panel data sets, however, the incorporation of heterogeneity in both the observations and the coefficient estimates have not been done before. These characteristics are, nonetheless, desirable as in practice it is possible that not all individuals are affected in the same way by some event. An empirical application concerning population growth relating to the rate of natural increase and international migrant stock illustrates this property.

Keywords Structural breaks · Panel Data · Heterogeneity · Grouped Pattern · Population Growth

Contents

1 Introduction

When entering the 21st century, the world has reached a world population milestone of 6 billion people. Fast-forwarding to this very moment, the population has grown to 7.6 billion people already (Worldometers). Since the Industrial Revolution in the 18th century, the world has experienced a rapid exponential growth in terms of population (see Figure A.1). However, only in the past century, there has been an extraordinarily growth, with an increase from 1.65 billion people in 1900 to over 6 billion people in the year 2000. During this century, a period of high rate in population growth of over 2% per year was experienced, which is characterized as 'rapid growth' (McNicoll,1984). Collecting these observations of the population over time from different countries is an example of a panel data set. These panel data sets, which captures observations of multiple events obtained over a period of time from a specific set of individual observations, captures ample information. Therefore, the use of this type of data has been widely used. It is relevant in many specializations such as in the field of economics or finance, where stock prices are measured over time. But also in the medical, quality control and seismologic area the use of panel data is present. Forecasts can be made based on these past observations using regression analysis. For example, in seismology, it is useful to predict observations as it can detect earthquakes in an early stage to limit its consequences. However, we ought to be careful when applying such regression methods in order to obtain relevant statistics or characteristics about the data. Often, the heterogeneous nature of these data are not taken into account, which can lead to erroneous models. Hence, we capture this heterogeneity by considering two different aspects of it. Firstly, we consider that one aspect of heterogeneity might be due to the phenomena of structural changes. When analyzing the relationship between certain (economic) variables, structural changes could occur which is caused by certain events. For example, a decrease in employment could be a consequence of an event as the European Debt Crisis in 2008. The change in the relationship between these variables can be modeled by structural breaks in the parameters of the panel data model. Multiple studies concerning detection of structural changes in data has been conducted in the past. Davis et. al. (1995) investigated changes in parameter values by looking at the level shift of the autoregressive model or change in auto-covariance structure. Aue (2009) develops an asymptotic test which assesses the structural stability of (cross-)volatilities for multivariate nonlinear time series. Other change-point tests such as historical and sequential, single and multiple break point tests have been discussed by Andreou (2009) illustrated with financial time series. However, existing literature about break detection techniques only consider common structural breaks all the observations, while, in practice, it might be that not all individuals experience the same breaks at the same time. Therefore, we consider another aspect of heterogeneity, which is heterogeneity amongst the different observations.

When performing a regression analysis on panel data, it is assumed that all individuals share the same slope coefficients. However, in practice, this might not be the case as not all individuals or countries are affected in the same way by the same event. For example, the European Debt

Crisis in 2008 has affected countries within Europe far more severe than countries in Asia. Thus, the structural breaks for these individuals or countries can differ in time and size. Therefore, it is crucial to jointly regard heterogeneity and structural breaks, otherwise, it can lead to false breakpoints and inconsistent estimations of the coefficients. However, on the other hand, it is not efficient to consider each individual separately, because by doing so, the model does not make use of cross-sectional information and, therefore, does not capture any common pattern among these individuals. Hence, using a grouped pattern to model individual heterogeneity is more effective as the coefficients of the model are estimated more efficiently. Furthermore, a grouped pattern gives more understanding about the correspondence between individuals.

Therefore, we consider the estimation method from Okui & Wang (2018), which introduces an estimation procedure that detects heterogeneous structural breaks. This procedure is called Grouped Adaptive Group Fused Lasso (GAGFL) and is a combination of the Grouped Fixed Effects (GFE) method described by Bonhomme & Manresa (2015a) and the Adaptive Group Fused Lasso (AGFL) by Qian & Su (2016a). Okui & Wang (2018) consider a linear panel data model with time-varying and heterogeneous coefficients, where individual heterogeneity is measured through a grouped pattern. The GFE method is used to determine the group memberships within the data, following a K-Means algorithm approach. At the same time, AGFL is used to estimate the coefficients and the structural breaks within each group by minimizing a least squares objective function. This procedure makes sure that the number of breaks can be determined together with consistently estimating the break points, the latent group memberships and the regression coefficients. Our aim is to revisit this method, in particular its Monte Carlo simulation results, to examine whether similar results can be obtained again.

Additionally, the proposed method by Okui & Wang (2018) is tested on an empirical application concerning the determinants of population growth across 118 different countries. The observations of each country are taken over a time series with 5-year frequency from 1965 up and until 2015. We measure the grouped pattern over different numbers of groups G, where G ranges from 1 to 5 groups. We found that the optimal number of groups is in fact $G = 1$ based on the Bayesian Information Criterion. However, the first breaks are detected for $G = 3$. Therefore, we report the GAGFL estimator results for the different groups $G = 1, \ldots, 3$. We found that our hypothesis of grouping countries based on their economical status have been (partly) rejected as some groups contain a combination of developed and developing countries. However, sensible arguments for this unexpected combination are provided.

The outline of the rest of the paper is as follows. Section 2 provides additional information which stresses the exponential growth in population during the last decades and describes the literature related to the GAGFL method. The method itself is explained in detail in Section 3. Then, by means of a Monte Carlo simulation, the finite sample performance of the GAGFL estimator is measured and reported in Section 4. In Section 5, GAGFL is applied to an empirical application concerning population growth. And lastly, we end with some concluding remarks and discussion points in Section 6.

2 Literature Review

As mentioned before, break detection methods are widely used and hence relevant in different types of specializations, such as applications concerning quality control and dynamical systems, which are often associated within fields of engineering. Taylor (2000) developed the change-point analysis method which is capable of detecting various changes in data and thereby provides the confidence level of which this change has occurred and also the confidence interval of when it occurred. Another application of break detection method is used in the field of seismology. Here, seismic signals are detected by considering significant changes in the characteristics of a time series such as power spectrum and variance (Kitagawa et. al., 2007). Additionally, it is also desirable to detect changes in patterns in the medical field. Brain wave patterns, or Electroencephalograms (EEG's), are examined to see what effects, for example, epileptic seizures have on the electrical activity of a local area of the brain and how this spreads through the rest of the brain. Ombao et. al. (2001) regards periods of patterns which fall out of the normal brain wave pattern as non-stationary. These periods of non-stationarity can be found by dividing the time series data into segments which are stationary, also called piecewise stationary process.

Research regarding individual heterogeneity has been done, mainly by considering heterogeneity in the slope coefficients. Several tests have been developed to check for differences in the slope coefficient estimators over cross-sectional observations. Juhl & Jugovvsky (2014) propose a test for slope heterogeneity in fixed effects models by extending the test described by Pesaran & Yamagata (2008), which is based on the conditional Gaussian likelihood function. Baltagi (2008) has examined ways to extend the Chow test (Chow, 1960) to test for slope coefficient heterogeneity against "poolability" of the data. Su et. al. (2016) studied the identification and estimation of the latent grouped pattern in panel data by using penalized techniques. The classifier-Lasso (C-Lasso) is used in order to simultaneously classify the observations and estimate the coefficients. Bonhomme & Manresa (2015) introduce a "grouped fixed-effects" estimator which estimates the parameters of the model by minimizing a least squares criterion while taking into account the underlying grouped pattern of the data.

When considering the methods of previous research mentioned above regarding break detection, these do not take into account the heterogeneity of panel data. It is assumed that all cross-sectional observations experience the same number of breaks at the same points in time. A similar reversed problem holds for clustering approaches to account for individual heterogeneity. When clustering the cross-sectional observations, each observation within a group share the same coefficients, however, these coefficients do not contain structural breaks. It is important to take into account the heterogeneity of the observations when estimating the slope coefficients, otherwise, neglecting the latent grouped pattern can lead to inconsistent estimation of the coefficients (Hsiao, 2014). Hence, the GAGFL method provides added value to existing literature by considering individual heterogeneity together with structural breaks, such that heterogeneous structural breaks can occur to observations within each group and can differ in time and size throughout different groups.

We aim to apply this proposed method to a societal empirical application which concerns the growth of the population all over the world. As time passes, the number of people on Earth increases at an alarming pace. In more than half a century, from 1960 until 2017, the world population grew from 3 billion to 7.6 billion, according to the United Nations, which is more than twice as large. McNicoll (1984) states that countries that experience 'rapid growth' are the countries which have an annual increase of around 2 percent or more in terms of their population. This increment of 2 percent per year results in a doubling of size in 35 years. Hence, rapid-growth countries double their population size in merely a generation. Furthermore, analyses of fertility and mortality patterns in the past suggest that this growth pattern will remain similar in the future (Bloom & Freeman, 1986). Humanity's impact on the natural environment on Earth is one of the greatest consequences of this change in world population. According to Ehrlich & Holdren (1971), a negative disproportionate impact on the Earth's environment will be one of the effects of population growth. This is a result of each human being partly accountable for the shrinkage of ecological systems due to its need of agriculture. Additionally, each person will make use of either renewable or nonrenewable energy sources as a result of technological advances which leads to exhaustion of the Earth. Kremer (1993) argues that larger populations will experience a faster technological change and population growth. Hence, the effects of technological advancement, which results in more usage of (non)renewable energy sources, and population growth are mutually reinforcing. Furthermore, there exists a severe food problem according to Pimentel et. al. (1997), since there is an astonishing amount of 1 to 2 billion people being malnourished. As the world population increases, this issue will only become more severe.

Since there seems to be an ever-increasing growth in world population, and considering its negative impact on the Earth's environment, it might be interesting to investigate whether there are structural breaks in this growth. These breaks in time could give insight in which events may have caused the growth to temporarily stabilize. This information could help analyzing the possible solutions to the exponential growth of the past century. Furthermore, by clustering the countries, it can be easily seen which group of countries experience similar population growth such that these countries can be categorized and targeted with a suitable policy. We expect that countries will be clustered together based on their economical status, since there is an association between the rate of population growth and the level of economic development of a country (Coale & Hoover, 1958).

3 Methodology

We use a linear panel data model with time-varying and heterogeneous coefficients, where the variance in time is captured in the coefficients in the form of structural breaks and heterogeneity is modeled via a grouped pattern. The panel data is described by a scalar dependent variable y_{it} and the regressors x_{it}, which is a vector of size k x 1, also represented as $\{\{y_{it}, x_{it}\}_{i=1}^{T}\}_{i=1}^{N}$, where $i = 1, \ldots, N$ and $t = 1, \ldots, T$ denote he observational unit and time, respectively. The linear panel data model can then be described as follows:

$$y_{it} = x_{it}^{'}\beta_{g_i,t} + \epsilon_{it}, \qquad i = 1, \ldots, N, \qquad t = 1, \ldots, N \tag{1}$$

where $E(\epsilon_{it}) = 0$. The coefficient $\beta_{i,t}$ is individual- and time-specific, however, as mentioned before, we put a restriction on its heterogeneity by assuming that it is group-specific for the sake of estimation and interpretation of the model. The group memberships are then defined as $g_i \in \mathbb{G}$ where g_i denotes to which group individual i belongs and $\mathbb{G} = \{1, \ldots, G\}$ denotes the set of groups. However, g_i is unknown and hence, has to be estimated. Furthermore, for each group g_i, the coefficients $\{\beta_{g,t}, \ldots, \beta_{g,T}\}$ portray the structural breaks. The structural changes are group-specific and we denote the number of breaks for each group as m_g and the set of break dates as $\tau_{m_g,g} = \{T_{g,1}, \ldots, T_{g,m_g}\}$. The coefficient $\beta_{g,t}$ will remain constant throughout the period between two break dates. This can be denoted as:

$$\beta_{g,t} = \alpha_{g,j}, \qquad \text{if} \qquad T_{g,j-1} \leq t < T_{g,j} \tag{2}$$

where $\alpha_{g,j}$ with $j = 1, \ldots, m_g$ is the coefficient-value until break date j and $T_{g,0} = 1$ and $T_{g,m_{g+1}} = T + 1$.

As mentioned before, the coefficients $\beta_{g,t}$, $\forall g \in \mathbb{G}$ and $\forall t \in \{1, \ldots, T\}$ and group membership variable g_i, $\forall i \in \{1, \ldots, N\}$ have yet to be estimated. We use the estimation method GAGFL, which is a combination of GFE by Bonhomme & Manresa (2015a) and AGFL by Qian & Su (2016a), where GFE is a clustering method and AGFL is a break detection method. In order to do so, we introduce the following notation: $\beta = (\beta_{1,1}^{'}, \ldots, \beta_{1,T}^{'}, \beta_{2,1}^{'}, \ldots, \beta_{G,T}^{'})$ and $\gamma = \{g_1, \ldots, g_N\}$. Denoting $\mathcal{B} \subset \mathbb{R}^k$ as the parameter space for each $\beta_{g,t}$, then the parameter space for β is $\mathcal{B}^{GT}$. As $\gamma \in \mathbb{G}^N$, the parameter space for γ is thus $\mathbb{G}^N$. Hence, we need to estimate (β, γ). We do this by solving the following penalized objective function:

$$(\hat{\beta}, \hat{\gamma}) = \underset{(\beta,\gamma) \in \mathcal{B}^{GT} \text{x} \mathbb{G}^N}{\text{argmin}} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - x_{it}^{'}\beta_{g_i,t})^2 + \lambda \sum_{g \in \mathbb{G}} \sum_{t=2}^{T} \dot{w}_{g,t} ||\beta_{g,t} - \beta_{g,t-1}|| \tag{3}$$

where the second term is called the adaptive Lasso penalty. This Lasso penalty makes use of a tuning parameter λ. Following Qian & Su (2016a), this tuning parameter is selected by minimizing the following information criterion:

$$IC(\lambda) = \frac{1}{NT} \sum_{j=1}^{m+1} \sum_{t=T_j-1+1}^{T_j} \sum_{i=1}^{N} (y_{it} - x_{it}^{'}\hat{\alpha}_{g_i,j})^2 + \rho_{NT} k(m_\lambda + 1) \tag{4}$$

where $\hat{\alpha}_{g_i,j}$, m_λ and ρ_{NT} are the post-Lasso coefficient estimates of g and j, the number of breaks and a factor determining the level of penalty on number of breaks, respectively. In line with Qian & Su (2016a), we choose $c = 0.05$ for $\rho_{NT} = c\ \ln(NT)/\sqrt{NT}$. Furthermore, the weights $\dot{w}_{g,t}$ used in Equation 3 are defined as $\dot{w}_{g,t} = ||\dot{\beta}_{g,t} - \dot{\beta}_{g,t-1}||^{-\kappa}$, with κ being an individual-specific constant and $\dot{\beta}$ a prior estimation for β. The preliminary estimates $\dot{\beta}$ and $\dot{\gamma}$ are determined by minimizing the function:

$$(\dot{\beta}, \dot{\gamma}) = \underset{(\beta,\gamma)\in\mathcal{B}^{GT}\text{x}\mathbb{G}^N}{\text{argmin}} \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - x_{it}^{'}\beta_{g_i,t})^2 \tag{5}$$

Note that in case $\beta_{g,t} - \beta_{g,t-1} = 0$, then $\dot{\beta}_{g,t} - \dot{\beta}_{g,t-1} \rightarrow 0$ and $\dot{w}_{g,t} \rightarrow \infty$ which translates to a large penalty term. When $\hat{\beta}_{g,t} - \hat{\beta}_{g,t-1} \neq 0$, then this period represents an estimation of a break date. Then, the number of structural breaks for group g can be determined by simply counting the periods of break dates, which can be described by the cardinality of the set of estimated break dates $\hat{\mathcal{T}}_g = \{t \in \{2, \ldots, T\} \mid \hat{\beta}_{g,t} - \hat{\beta}_{g,t-1} \neq 0\}$.

In order to minimize the penalized objective function as in Equation 3 we make use of an iterative procedure where we estimate γ by minimizing the sum of squared errors of each unit. Hence, we estimate g_i, which is the group membership of individual i, by checking which group membership assignment results in the smallest sum of squared residuals for individual i. Thereafter, if the clustering structure is obtained, we estimate the β coefficients by applying the break detection method AGFL to the each group. The algorithm is described below:

Algorithm 1. *Set $\gamma^{(0)}$ as the initial GFE estimate of grouping $\dot{\gamma}$, and $s = 0$.*
Step 1: *For given $\gamma^{(s)}$, compute:*

$$\beta^{(s)} = \underset{\beta\in\mathcal{B}^{GT}}{\text{argmin}} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - x_{it}^{'}\beta_{g_i^{(s)},t})^2 + \lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}||\beta_{g,t} - \beta_{g,t-1}|| \tag{6}$$

Step 2: *Compute for all $i \in \{1, \ldots, N\}$*

$$g_i^{(s+1)} = \underset{g\in\mathcal{G}}{\text{argmin}} \sum_{t=1}^{T}(y_{it} - x_{it}^{'}\beta_{g,t}^{(s)})^2$$

Step 3: *Set $s = s + 1$. Go to Step 1 until numerical convergence.*

The above algorithm can be described as first applying the AGFL method by Qian & Su (2016a) to each estimated group. Subsequently, the group membership is updated by minimizing the objective function, which is the sum of squared residuals. These steps are repeated until numerical convergence is reached. Choosing $(\dot{\beta}, \dot{\gamma})$ as initial values, the algorithm will converge quickly as these initial values are shown to be consistent by Okui & Wang (2018). However, since there is no solid initialization for GFE as described in Equation 5, because there is no certainty that clustering method GFE will result in a global minimum, a large number of random initial values are drawn in order to estimate $\dot{\gamma}$.

4 Monte Carlo Simulation

This section tests the GAGFL method by means of a simulation. The data-generating process, evaluation criteria and the results of the analysis are described in the next subsections.

4.1 Data Generation Process and Procedures

We generate data for the simulation using the following data-generating process (DGP) with regression model:

$$y_{it} = x_{it}\beta_{g_i,t} + \epsilon_{it}$$

where the regressors are $x_{it} \overset{\text{i.i.d.}}{\sim} N(0,1)$ and $\epsilon_{it} \overset{\text{i.i.d.}}{\sim} N(0,\sigma_\epsilon^2)$, with $\sigma_\epsilon = (0.5, 0.75)$. As mentioned before, we consider $\{\{y_{it}, xi_{it}\}_{t=1}^T\}_{i=1}^N$, hence there are N cross-sectional observations with each a time series of length T. The cross-sectional observations are divided into three groups, N_j with $j = 1,2,3$, where N_j denotes the number of units per group j. Consistently, the units in the three groups N_1, N_2 and N_3 must add up to the total number of units N. Furthermore, the total number of observations are divided among the three groups in a ratio of $N_1 : N_2 : N_3 = 0.3 : 0.3 : 0.4$. Groups N_1 and N_2 incorporate two structural breaks in the slope coefficients, both at two different points in time. Namely, group N_1 has breaks at time points $\lfloor T/2 \rfloor$ and $\lfloor 5T/6 \rfloor$ and group N_2 at the points $\lfloor T/3 \rfloor$ and $\lfloor 5T/6 \rfloor$, where $\lfloor \cdot \rfloor$ is the rounded down value. On the other hand, group N_3 contains no breaks and, thus, exhibits a constant slope coefficient throughout the whole time series length T. The slope coefficients are defined as follows:

$$\beta_{1,t} = \begin{cases} 1 & \text{if } 1 \leq t < \lfloor T/2 \rfloor \\ 2 & \text{if } \lfloor T/2 \rfloor \leq t < \lfloor 5T/6 \rfloor \\ 3 & \text{if } \lfloor 5T/6 \rfloor \leq t \leq T \end{cases} \quad , \qquad \beta_{2,t} = \begin{cases} 3 & \text{if } 1 \leq t < \lfloor T/3 \rfloor \\ 4 & \text{if } \lfloor T/3 \rfloor \leq t < \lfloor 5T/6 \rfloor \\ 5 & \text{if } \lfloor 5T/6 \rfloor \leq t \leq T \end{cases}$$

and

$$\beta_{3,t} = 1.5 \quad \text{for all} \quad 1 \leq t \leq T$$

For our simulation, the number of cross-sectional observations can take two different values $N = (50, 100)$ and the length of the time series can take three different values $T = (10, 20, 40)$. Furthermore, recall that the standard deviation of the errors in the DGP can also take two distinct values $\sigma_\epsilon = (0.5, 0.75)$. Hence, with these variables, 12 different combinations and thus, 12 different datasets can be made.

4.2 Evaluation Criteria

In order to analyze the correctness of the method, several evaluation criteria must be met. We measured the performance of GAGFL based on its ability to correctly cluster the observations and accurately estimate the break dates and slope coefficients. Due to time constraints, a total of 100 replications are run instead of 1000 replications as described in Okui & Wang (2018). Nonetheless, similar results have been achieved, as shown in Section 4.3.

4.2.1 Clustering Criteria

The accuracy of clustering the observations is measured by computing the average of the misclassification frequency (MF) across 100 replications. MF is defined as the number of misclassified observations divided by the total number of observations:

$$MF = \frac{1}{N}\sum_{i=1}^{N} I(\hat{g}_i \neq g_i^0)$$

where $I(\cdot)$ represents an indicator function and $\hat{g}_i$ and g_i^0 are the estimated and true group membership of unit i, respectively.

4.2.2 Break Date Criteria

The break estimations are assessed by two different criteria. Firstly, we measure whether GAGFL can accurately estimate the number of breaks in each group. This can be done by computing the average frequency of estimating the number of breaks correctly in each group. Furthermore, we also keep track of the average estimated number of breaks per group. Secondly, the correctness of the break date estimates are evaluated based on their Hausdorff error, as done by Qian & Su (2016). The Hausdorff error is based on the Hausdorff distance (HD), which is defined as follows:

$$HD(\hat{T}^0_{g,\hat{m}}, T^0_{g,m^0}) \equiv \max\{\mathcal{D}(\hat{T}^0_{g,\hat{m}}, T^0_{g,m^0}), \mathcal{D}(T^0_{g,m^0}, \hat{T}_{g,\hat{m}})\}$$

Here, $\mathcal{D}(A, B) \equiv \sup_{b\in B}\inf_{a\in A}|a - b|$ for any set A and B and $\hat{T}^0_{g,\hat{m}}$ and T^0_{g,m^0} are the estimated break dates and true break dates for each group, respectively. Then, the Hausdorff error is reported as the average of $HD(\hat{T}^0_{g,\hat{m}}, T^0_{g,m^0})/T$ across the 100 replications.

4.2.3 Slope Coefficients Criteria

Lastly, the accuracy of the slope coefficient estimates are assessed by their root mean squared error (RMSE) and the coverage probability of the two-sided nominal 95% confidence interval. The RMSE is computed as:

$$\text{RMSE}(\hat{\beta}_{it}) = \sqrt{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{\beta}_{it} - \beta_{it})^2}$$

where $\hat{\beta}_{it}$ is the estimated slope coefficient and β_{it} the true slope coefficient. Additionally, in order to compute the coverage probability, we need $\hat{\sigma}_{\beta,it}$, which is the estimated standard deviation of $\hat{\beta}_{it}$. Then, the coverage probability is defined as follows:

$$\text{Coverage}(\hat{\beta}_{it}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} I(\hat{\beta}_{it} - 1.96\hat{\sigma}_{\beta,it} \leq \beta_{it} \leq \hat{\beta}_{it} + 1.96\hat{\sigma}_{\beta,it})$$

4.3 Results Monte Carlo Simulation

We replicate the results displayed in the first parts of Table 2 through 6 in the Technical Appendix from Okui & Wang (2018), which concerns the examination of data generated by DGP 1. Due to time constraints, only 100 replications instead of 1000 replications are used. However, similar results as in Okui & Wang (2018) are achieved by this limited amount of replications. We evaluate the GAGFL method by testing its ability to correctly determine the number of breaks, break date estimates and coefficient estimates in each group and its ability to accurately assign each data point to its original cluster.

4.3.1 Clustering Accuracy

Table 1 describes the average misclassification frequency for each combination of N = (50,100), T = (10, 20, 40) and σ_ϵ = (0.5, 0.75) across the replications. We notice that for $\sigma_\epsilon = 0.5$, the misclustering rate is very close to zero, only for T = 10, it has a small deviation from zero on the fifth decimal place. Furthermore, the rate for N = 100 is smaller than for N = 50, which means more observations lead to more correct clustering. However, the increase in T has the biggest impact on the rate as for both $\sigma_\epsilon = 0.5$ and $\sigma_\epsilon = 0.75$ holds that the misclassification frequency exponentially decreases as T increases. Therefore, having a longer times series length per individual is more favorable than having more individuals in the data, because the misclustering rate decreases faster when T increases than when N increases. Hence, the GAGFL method can reliably retrieve the original group memberships of the observations, especially when T is large.

	N = 50			N = 100		
	T = 10	T = 20	T = 40	T = 10	T = 20	T = 40
sigma = 0.5	0.00004	0.0000	0.0000	0.00006	0.0000	0.0000
sigma = 0.75	0.0504	0.0102	0.0004	0.0288	0.0016	0.0000

Table 1: Average Misclustering Rate

4.3.2 Break Date Accuracy

Table 2 until Table 4 concerns the assessment of the break dates. Specifically, Table 2 displays the frequency of correct estimation of number of breaks corresponding to each group. Table 3 shows the average estimated number of breaks for each group. And from Table 4 can be deduced whether the break date estimates are correct by looking at the Hausdorff error.

As can be seen in Table 2, the number of breaks are 100% correctly estimated when using $\sigma_\epsilon = 0.5$ in the DGP for all combinations of N and T. However, there is a considerable change when $\sigma_\epsilon = 0.75$ is used. In the worst case, an average frequency of only 60% of correct estimation of number of breaks is measured for N = 50 and T = 10. Nonetheless, the rate steadily improves when N and T increase. Hence, we confirm the findings of Okui & Wang (2018) that the GAGFL method (almost) perfectly detects the correct number of breaks in case of moderate size errors ($\sigma_\epsilon = 0.5$).

		N = 50			N = 100		
	Group (True break num.)	T = 10	T = 20	T = 40	T = 10	T = 20	T = 40
σ_ϵ = 0.5	G1 ($m^0_{1,0} = 2$)	1.000	1.000	1.000	1.000	1.000	1.000
	G2 ($m^0_{2,0} = 2$)	1.000	1.000	1.000	1.000	1.000	1.000
	G3 ($m^0_{3,0} = 0$)	1.000	1.000	1.000	1.000	1.000	1.000
σ_ϵ = 0.75	G1 ($m^0_{1,0} = 2$)	0.580	0.720	0.890	0.820	0.930	0.990
	G2 ($m^0_{2,0} = 2$)	0.640	0.730	0.900	0.840	0.960	0.990
	G3 ($m^0_{3,0} = 0$)	0.580	0.890	1.000	0.940	0.950	1.000

Table 2: Average Frequency of Correct Estimation of the Number of Breaks

Table 3 shows the average estimated number of breaks for each group. The true number of breaks for each group ($m^0_{g,0}$) is 2, 2 and 0, respectively. As already noted from Table 2: the number of breaks are 100% correctly estimated for $\sigma_\epsilon = 0.5$, which implies that the average number of breaks are also correctly estimated. This is indeed the case, as all combinations of N and T for $\sigma_\epsilon = 0.5$ have the correct number of groups: 2, 2, 0. In case of $\sigma_\epsilon = 0.75$, the average number of groups are all slightly over-estimated, but are approaching the true number of groups as N and T increase.

		N = 50			N = 100		
	Group (True break num.)	T = 10	T = 20	T = 40	T = 10	T = 20	T = 40
σ_ϵ = 0.5	G1 ($m^0_{1,0} = 2$)	2.000	2.000	2.000	2.000	2.000	2.000
	G2 ($m^0_{2,0} = 2$)	2.000	2.000	2.000	2.000	2.000	2.000
	G3 ($m^0_{3,0} = 0$)	0.000	0.000	0.000	0.000	0.000	0.000
σ_ϵ = 0.75	G1 ($m^0_{1,0} = 2$)	2.550	2.370	2.120	2.170	2.070	2.010
	G2 ($m^0_{2,0} = 2$)	2.510	2.400	2.110	2.190	2.040	2.010
	G3 ($m^0_{3,0} = 0$)	0.72	0.140	0.000	0.080	0.050	0.000

Table 3: Average Estimated Number of Breaks

After examining whether the number of breaks are correctly estimated, we check whether the break date estimates are also correct. The results can be found in Table 4, which displays the Hausdorff errors of the break date estimates per group. Naturally, the smaller the errors, the more favorable. For $\sigma_\epsilon = 0.5$, the reported errors are all zero. Combined with the facts that the number of breaks and frequency of correct number of breaks are also optimal, we can imply that for $\sigma_\epsilon = 0.5$ the break date estimates are correct and are identical to the true break dates in our DGP. Even when $\sigma_\epsilon = 0.75$, the errors are still close to zero and become smaller when N and T increase. Hence, we can again confirm the findings of Okui & Wang (2018) that the GAGFL method can identify break points accurately.

		N = 50			N = 100		
	Group (True break num.)	T = 10	T = 20	T = 40	T = 10	T = 20	T = 40
$\sigma_\epsilon = 0.5$	G1 ($m^0_{1,0} = 2$)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	G2 ($m^0_{2,0} = 2$)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	G3 ($m^0_{3,0} = 0$)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\sigma_\epsilon = 0.75$	G1 ($m^0_{1,0} = 2$)	0.0550	0.0355	0.0140	0.0220	0.0105	0.0013
	G2 ($m^0_{2,0} = 2$)	0.0550	0.0345	0.0113	0.0220	0.0040	0.0001
	G3 ($m^0_{3,0} = 0$)	0.1460	0.0305	0.0000	0.0230	0.0100	0.0000

Table 4: Hausdorff Error of Break Date Estimates

4.3.3 Slope Coefficient Accuracy

Lastly, Table 5 and 6 examine the accuracy of the coefficient estimates by their root mean squared errors (RMSE) and the coverage probability, where the desired results are a small RMSE and a coverage probability of around 95% as this probability encloses the two-sided nominal 95% confidence interval.

Table 5 displays the RMSE's corresponding to each combination of N, T and σ_ϵ. For $\sigma_\epsilon = 0.5$, the RMSE decreases with an average of approximately 30% as T doubles. For $\sigma_\epsilon = 0.75$, the RMSE decreases slighty faster with an average of 42%. However, in both cases, the RMSE for T = 10 were already quite small.

	N = 50			N = 100		
	T = 10	T = 20	T = 40	T = 10	T = 20	T = 40
$\sigma_\epsilon = 0.5$	0.0370	0.0248	0.0174	0.0262	0.0166	0.0134
$\sigma_\epsilon = 0.75$	0.1647	0.0869	0.0481	0.0781	0.0463	0.0296

Table 5: Root Mean Squared Error of Coefficient Estimates

Table 6 displays the coverage probability of the coefficient estimates. We notice for both $\sigma_\epsilon = 0.5$ and $\sigma_\epsilon = 0.75$ and for N = 50 that, as T increases, the coverage probability approaches the value around 95%. We can confirm that, as the sample size increases, the RMSE becomes smaller and the coverage probabilities are improved, which is in line with the findings of Okui & Wang (2018). A small RMSE and adequate coverage probability implies that the coefficients from the GAGFL method are properly estimated.

	N = 50			N = 100		
	T = 10	T = 20	T = 40	T = 10	T = 20	T = 40
$\sigma_\epsilon = 0.5$	0.9281	0.9464	0.9466	0.9596	0.9568	0.9618
$\sigma_\epsilon = 0.75$	0.9379	0.9606	0.9602	0.9689	0.9516	0.9522

Table 6: Coverage Probability of Coefficient Estimates

5 Empirical Application

We propose to use the GAGFL method by Okui & Wang (2018) on a new empirical application which studies the relationship between the rate of population growth and rate of natural increase in population and international migrant stock.

5.1 Data

A balanced panel data set is used, extracted from the World Bank, concerning crude birth and death rates, international migrant stock, total amount of population per country and population growth rate of 118 countries spread across the world. These data contain 11 time periods with 5-year frequency from 1965 up and until 2015. The birth rate indicates the amount of live births throughout the year, per 1,000 people, estimated at midyear. Similarly, the death rate indicates the number of deaths throughout the year, per 1,000 population, also estimated at midyear. By subtracting the crude death rate from the crude birth rate, the rate of natural increase (RNI) can be calculated. RNI indicates the rate of change in population in absence of migration. Since the birth rate dominates the death rate for most countries, this results in a positive RNI with an average of 19.5 (see Figure A.2, which depicts the average RNI over the years 1960-2015 with a 95% confidence interval). The international migrant stock equals the amount of people which is born in another country than which they currently live in, including refugees. The total population is defined as the total count of people in the country, neglecting their citizenship or legal status. By dividing the international migrant stock by the total population per country, we obtained the percentage of international migrant stock (MS) relative to the population, which has an average of 0.08% (see Figure A.3). The annual population growth rate of year t equals the exponential growth rate of the population from year t-1 until year t, estimated at midyear. This is expressed as a percentage with an average rate of 1.96% per year (see Figure A.4).

5.2 Method/Regression

Our aim is to regress the rate of natural increase (RNI) and rate of international migrant stock (MS) on the rate of population growth (PG). Therefore, we consider a fixed effects model by including an additive country-specific, time-invariant fixed effect μ_i as described in the model below:

$$PG_{it} = \mu_i + \theta_{1,g_i,t} RNI_{it} + \theta_{2,g_i,t} MS_{it} + \epsilon_{it}$$

However, GAGFL tend to give more accurate results when the data is stationary. Hence, we eliminate the trend by using first-differenced data, following Okui & Wang (2018).

$$\Delta PG_{it} = \theta'_{1,g_i,t} \Delta RNI_{it} + \theta'_{2,g_i,t} \Delta MS_{it} + \Delta\epsilon_{it}$$

Moreover, to ensure similar variation for grouping, the variables are standardized by subtracting the mean from the observation and divided by its standard deviation: $z_{it} = \frac{x_{it}-\mu_i}{\sigma_i}$. The coefficients are estimated by applying GAGFL on the transformed data, which looks as follows:

$$\Delta PG_{it} = \theta''_{1,g_i,t} \Delta RNI_{it} + \theta''_{2,g_i,t} \Delta MS_{it} + \Delta\epsilon_{it} \tag{7}$$

We implement GAGFL on the data as described in Equation 7. Hereby, we choose the tuning parameter λ by searching in the interval of $\lambda_{min} = 2$ and $\lambda_{max} = 50$ with a grid size of 200. We examine the groups and slope coefficients for $G = 1, \dots, 5$ and use 10,000 starting values for initializing the group memberships of each observation. In order to determine the optimal number of groups, we follow Bonhomme & Manresa (2015a) by considering the Bayesian Information Criterion (BIC). This information criterion is defined as follows:

$$BIC(G) = \frac{1}{NT} \sum_{j=1}^{m+1} \sum_{t=Tj-1+1}^{T_j} \sum_{i=1}^{N} (y_{it} - x'_{it}\hat{\alpha}_{g_i,j})^2 + \hat{\sigma}^2 \frac{n_p(G)+N}{NT} lnNT \tag{8}$$

The optimal number of groups corresponds to the lowest corresponding BIC-value.

5.3 Results

When implementing the BIC, the lowest value corresponds to G=1 (see Table A.1). This means that, according to this criterion, the countries should not be split up into different number of groups. However, as can be seen from Table 7, there are zero breaks detected when assigning G=1. Therefore, for the sake of illustrating the GAGFL method, we also report and describe the results for G=2 and G=3, which correspond to the second- and third-lowest BIC-value.

Firstly, we examine the results of the regression as in Equation 7, assuming that the number of groups G is equal to 1, which is displayed in Table 7. The number in-between brackets indicate the standard error of the corresponding coefficient, n represents how many observations are within a particular group and regimes $1, \dots, 11$ correspond to the time periods from 1965 up and until 2015 with 5-year frequency. For Group 1.1, there are zero breaks estimated and the slope coefficient for the rate of natural increase (RNI) is larger than the slope coefficient for the international migrant stock (MS). This implies that, in general, a country's population mainly grows because of its rate of natural increase rather than due to immigration.

	Regime	1	2	3	4	5	6	7	8	9	10	11
Group 1.1	RNI	0.5657 (0.0441)										
n = 118	MS	0.0859 (0.0342)										

Table 7: Coefficient and Regime Estimates of G = 1

Secondly, we now assume that the true number of groups G is equal to 2. The results are shown in Table 8. The total of 118 countries are now split into a larger group, consisting of 83 countries, and a smaller group, incorporating 35 countries, denoted as Group 2.1 and 2.2, respectively. The smaller group is characterized by a slightly higher slope coefficient for international migrant stock (MS). Hence, this implies that for these countries, their population grew mainly because of immigration. Further analysis showed that almost half of the countries in this group is characterized as a 'high income'-country according to The World Bank Data (see Figure 9, which displays the division of countries per group based on income status in percentages). As can be seen in Figure A.5, this group contains relatively small and rich countries or regions such as Qatar, Hong Kong and Luxembourg. This finding affirms that, according to functionalist social theory, people tend to move from lower income-regions to higher income-regions as most migrants are portrayed as income-maximizing individuals (de Haas, 2011). For Group 2.1, the population mainly grew because of a high birth rate and low death rate, which results in a positive rate of natural increase (RNI). Contrary to Group 2.2, less than 20% of this group is labeled as 'high-income'-countries. There is, in general, an inverse correlation between the number of offspring and wealth. According to Balbo et. al. (2012), advanced societies are characterized by low-fertility levels, which is close to 1 child per woman. Hence, this could explain why the RNI is much higher for this group of countries. As this bigger group contains mostly Asian, South-American and North-African countries, such as Peru, Thailand, Indonesia, Morocco and India, which were economically developing countries in the period of 1965 - 2015. During this span of time, most of these countries were still industrializing and the income level and education of the population was not optimal yet. Furthermore, since developing countries offer little favorable living conditions, it is not encouraging for foreign people to migrate towards these places, which explains the relatively low slope coefficient for MS. Thus, these two groups are separated based on their economical status which is expressed through RNI and MS. However, no structural breaks in the parameters have been detected yet for either of these groups.

	Regime	1	2	3	4	5	6	7	8	9	10	11
Group 2.1	RNI	0.7655 (0.0314)										
n = 83	MS	0.0174 (0.0385)										
Group 2.2	RNI	0.1126 (0.0744)										
n = 35	MS	0.2117 (0.0611)										

Table 8: Coefficient and Regime Estimates of G = 2

		Income Status			
	%	High Income	Upper Middle Income	Lower Middle Income	Low Income
Groups	2.1	18.07	32.53	32.53	16.87
	2.2	48.57	20.00	14.29	17.14
	3.1	17.39	34.78	36.23	11.59
	3.2	40.00	28.00	16.00	16.00
	3.3	41.67	12.50	12.50	33.33

Table 9: Percentages of Countries Divided per Income Class

Proceeding with G = 3, we found that all the 69 countries from Group 3.1 are extracted from Group 2.1. The slope coefficient estimates from Group 3.1 are mostly similar to those of Group 2.1, where the coefficient estimate of RNI remains much larger than that of MS. However, the slope coefficient for MS becomes slightly smaller, such that it is now negative. The group membership of the countries can be seen in Figure A.6. Similarly to Group 2.1, less than 20% is considered as a high income-country for Group 3.1 as well, as can be seen from Table 9. The majority of Group 3.2, 23 out of 25 countries, originated from Group 2.2. Here, the two coefficient estimates for both variables RNI and MS are relatively similar, albeit that the slope coefficient estimate for MS is slightly higher than that of RNI. Furthermore, the high-income segment is the largest in this group by making up 40% of the group, followed by the upper middle income segment, which takes up 28%. The new group, Group 3.3, contains precisely half of its total of 24 countries from Group 2.1 and the other half from Group 2.2. Furthermore, the high-income segment takes up approximately 40% of the group, equivalently to Group 3.2. However, there are more low income-countries incorporated in this group (33.3%) in comparison to the number of low income-countries in Group 3.2 (only 16%). Group 3.3 is the first group where structural breaks are detected. The breaks are estimated between regimes 4 and 5 and regimes 8 and 9, which corresponds to the periods between 1980-1985 and 2000-2005, respectively. For the variable RNI, these structural breaks divides the time line into periods of declining population growth before 1980, then a sharp increase between 1985 and 2000 and from 2005 onwards, the rate of natural increase slightly decreases again. For the international migrant stock variable, the slope coefficients steadily decrease from 1965 up and until 2015, where the slope coefficient is negative from 2005 onwards. The latter indicates that, concerning the countries in Group 3.3, the inhabitants slowly emigrate to other foreign places. Group 3.3 contains relatively many low and high income-countries. Low income-countries, such as Uganda, Ethiopia and Congo, experience fluctuations in population growth which might be high due to lethal diseases, as these countries are known to have limited health care services and because of (civil) wars. For example, the Congo crisis, which lasted from 1960 until 1965, was a period of political conflicts and have caused many deaths in the country. However, this group also contains wealthier countries, such as, Belgium, Ireland and Luxembourg. The initial negative coefficient estimate for RNI and high estimate of MS might be due to post-war consequences of, for example, WWII and the spike in RNI and decline in MS, thereafter, can be devoted to a period of flourishing economies as during this time, people deem it to be safe and are satisfied with their living conditions such that they can settle down.

<table>
<tr><th></th><th>Regime</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th><th>10</th><th>11</th></tr>
<tr><td>Group 3.1</td><td>RNI</td><td colspan="11">0.8050 (0.0268)</td></tr>
<tr><td>n = 69</td><td>MS</td><td colspan="11">-0.0327 (0.0406)</td></tr>
<tr><td>Group 3.2</td><td>RNI</td><td colspan="11">0.1851 (0.0899)</td></tr>
<tr><td>n = 25</td><td>MS</td><td colspan="11">0.2540 (0.0678)</td></tr>
<tr><td>Group 3.3</td><td>RNI</td><td colspan="4">-0.0355 (0.1390)</td><td colspan="4">0.5265 (0.1255)</td><td colspan="3">0.4410 (0.1587)</td></tr>
<tr><td>n = 24</td><td>MS</td><td colspan="4">0.2789 (0.1566)</td><td colspan="4">0.2378 (0.0901)</td><td colspan="3">-0.1341 (0.1253)</td></tr>
</table>

Table 10: Coefficient and Regime Estimates of G = 3

6 Conclusion & Discussion

When analyzing panel data, it is important to take into account the heterogeneity in its observations along with heterogeneity in the slope coefficients. Events over time might cause structural breaks in the parameters which should be captured in the model. Failing to do so will cause incorrect coefficient estimates and, thus, an erroneous model. Furthermore, forming grouped patterns to express heterogeneity among individuals is crucial as well, since not all individuals might be affected in the same way by a certain event. By combining the GFE estimator for grouping individuals and AGFL for break detection, the GAGFL method can simultaneously capture this underlying grouped pattern as well as estimating slope coefficient and break date estimates.

We analyzed the GAGFL method of Okui & Wang (2018) by means of a Monte Carlo simulation to assess its performance in finite samples. GAGFL is proven to give more accurate results as the length of the time series T increases, rather than when the number of observation N increases. It is worth to mention that not the exact same simulation has been performed, as in this paper only 100 replications, instead of 1000, are used due to time constraints. Nonetheless, similar results have been found as in Okui & Wang (2018) when evaluating it by the criteria of clustering, break date and slope coefficient accuracy.

Applying GAGFL on an empirical application, which considers the relationship between the rate of population growth and its determinants, we found our hypothesis of grouping economically similar countries into the same group being (partly) rejected. However, sensible arguments concerning war outbreaks and demographic factors are provided to why this is the case. A point of discussion for this empirical application is the choice of tuning parameter λ. When experimenting with different intervals of λ, we found that when choosing a small lower bound for λ, for example $\lambda = 0.001$, this corresponds to having breaks at every point in time. When increasing the lower bound, this results in less number of breaks. Further research could be conducted concerning choosing the interval for this tuning parameter in order to justify the correctness in number of breaks. In addition, since the clustering approach is based on the k-means algorithm, which involves random initializations, it is highly subjective to its initial values. In this empirical application, 10,000 initial values are used in order to determine the group memberships, however, with every run, different group memberships, break dates, number of breaks and slope coefficient are measured when using $G > 3$. Hence, this could indicate that there might be many local minima. It might be interesting to explore this random initialization strategy and to replace it with more deterministic techniques, such as hierarchical clustering, in order to achieve replicable results for empirical applications.

References

Anderson, G.F. & Hussey, P.S. (2000). Population Aging: A Comparison Among Industrialized Countries. *Health Affairs*, Vol. 19, No. 3, pp. 191-203

Andreou, E. & Ghysels, E. (2009). Structural Breaks in Financial Time Series. In: Mikosch T., Kreiß JP., Davis R., Andersen T. (eds) Handbook of Financial Time Series. Springer, Berlin, Heidelberg

Aue, A,., Hörmann, S., Horváth, L. & Reimherr, M. (2009). Break Detection in the Covariance Structure of Multivariate Time Series Models. *The Annals of Statistics*, Vol. 37, No. 6B, pp. 4046–4087

Balbo, N., Billari, F.C. & Mills, M. (2012). Fertility in Advanced Societies: A Review of Research. *European Journal of Population*, Vol. 29, No. 1, pp. 1-38

Baltagi, B. H. (2008). *Econometric Analysis of Panel Data* (Fourth ed.). Chichester: John Wiley & Sons. ISBN 978-0-470-51886-1

Bloom, D.E., & Freeman, R.B. (1986). The Effects of Rapid Population Growth on Labor Supply and Employment in Developing Countries. *Population and Development Review*, Vol. 12, No. 3, pp. 381-414

Bonhomme, S., & Manresa, E. (2015a). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83:1147-1184.

Bonhomme, S., & Manresa, E. (2015b). Supplement to 'grouped patterns of heterogeneity in panel data'. *Econometrica Supplemental Material*, 83:1147-1184.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica* 28: 591–605 .

Coale, A.J. & Hoover, E.M. (1958). Population Growth and Economic Development in Low-Income Countries. A Case Study of India's Prospects. Princeton, N.J.: Princeton University Press

Davis, R., Huang, D., & Yao, Y. (1995). Testing for a Change in the Parameter Values and Order of an Autoregressive Model. *The Annals of Statistics*, 23(1), 282-304.

Ehrlich, P.R., & Holdren, J.P. (1971). Impact of Population Growth. *Science*. New Series, Vol. 171, No. 3977, pp. 1212-1217

Gibbs, D.N. (1991). *The Political Economy of Third World Intervention: Mines, Money, and U.S. Policy in the Congo Crisis*. Chicago and London: the University of Chicago Press.

Gilles, D. (2007, October 14). Kabul at War (1992-1996) : State, Ethnicity and Social Classes. *South Asia Multidisciplinary Academic Journal*. Retrieved June 22, 2018, from https://journals.openedition.org/samaj/212

de Haas, H. (2011). The determinants of international migration (Working Paper). International Migration Institute.

Hsiao, C. (2014). *Analysis of Panel Data* (Econometric Society Monographs). Cambridge: Cambridge University Press.

Juhl, T. & Lugovvsky, O. (2014). A Test for Slope Heterogeneity in Fixed Effects Models. *Econometric Reviews*, Vol. 33, No. 8, pp. 906-935

Kitagawa, G., Takanami, T. & Matsumoto, N. (2001). Signal Extraction Problems in Seismology. *International Statistical Review*, vol. 69, no. 1, pp. 129–152

Kremer, M. (1993). Population Growth and Technological Change: One Million B.C. to 1990. *The Quarterly Journal of Economics*, Vol. 108, No. 3, pp. 681-716

McNicoll, G. (1984). Consequences of Rapid Population Growth: An Overview and Assessment. *Population and Development Review* Vol. 10, No. 2, pp. 177-240

Okui, R., & Wang, W. (2018). Heterogeneous structural breaks in panel data models. *SSRN Electronic Journal*. doi:10.2139/ssrn.3031689

Ombao, H.C., Raz, J.A., Sachs, R. von, & Malow, B.A. (2001). Automatic Statistical Analysis of Bivariate Nonstationary Time Series. *Journal of the American Statistical Association*, vol. 96, no. 454, 2001, pp. 543–560

Pesaran, M. H. & Yamagata, T. (2008). Testing slope homogeneity in large panels. *Journal of Econometrics* 142: 50–93 .

Pimentel, D., Huang, X., Cordova, A., Pimental, M. (1997). Impact of Population Growth on Food Supplies and Environment. *Population and Environment: A journal of Interdisciplinary Studies*. Vol.19, No.1.

Qian, J. & Su, L. (2016a). Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso. *Journal of Econometrics*, 191:86-109.

Qian, J. & Su, L. (2016a). Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso. *Journal of Econometrics*, 191:86-109.

Shrestha, L.B. (2000). Population Aging In Developing Countries. *Health Affairs*, Vol. 19, No. 3, pp. 204-212.

Su, L., Shi, Z. & Phillips, P.C.B. (2014). Identifying Latent Structures in Panel Data. *Econometrica*, Vol. Vol. 84, No. 6, pp. 2215–2264

Taylor, Wayne A. (2000), Change-Point Analysis: A Powerful New Tool For Detecting Changes. Retrieved June 17, 2018, from *http://www.variation.com/cpa/tech/changepoint.html*

United Nations Population Division, Department of Economic and Social Affairs. *The World at Six Billion* (United States of America, 1999), available from http://www.un.org/esa/population/publications/sixbillion/sixbillion.htm

Worldometers (2018). Current world population. Retrieved May 17, 2018, from *http://www.worldometers.info/world-population/*.

Zack-Williams, A.B. (1999). Sierra Leone: The political economy of civil war, 1991-98. *Third World Quarterly*, 20:1, pp. 143-162

A Appendix

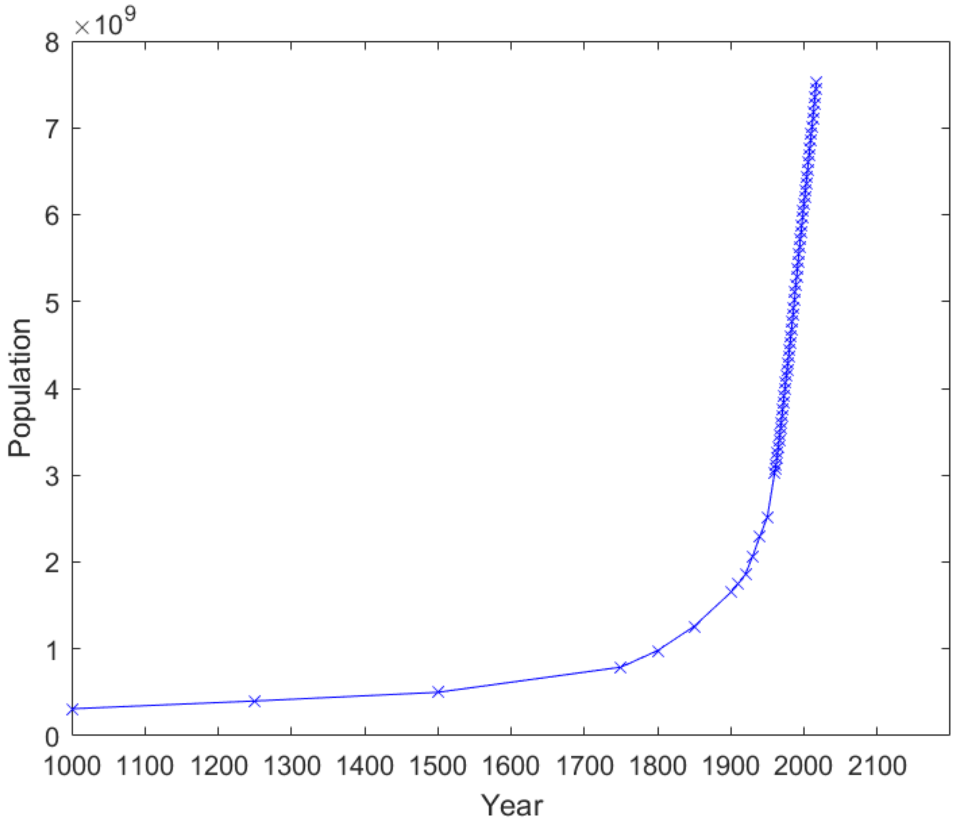

Figure A.1: Total World Population From the Year 1000 until 2017

Source: United Nations Population Division, The World Data Bank

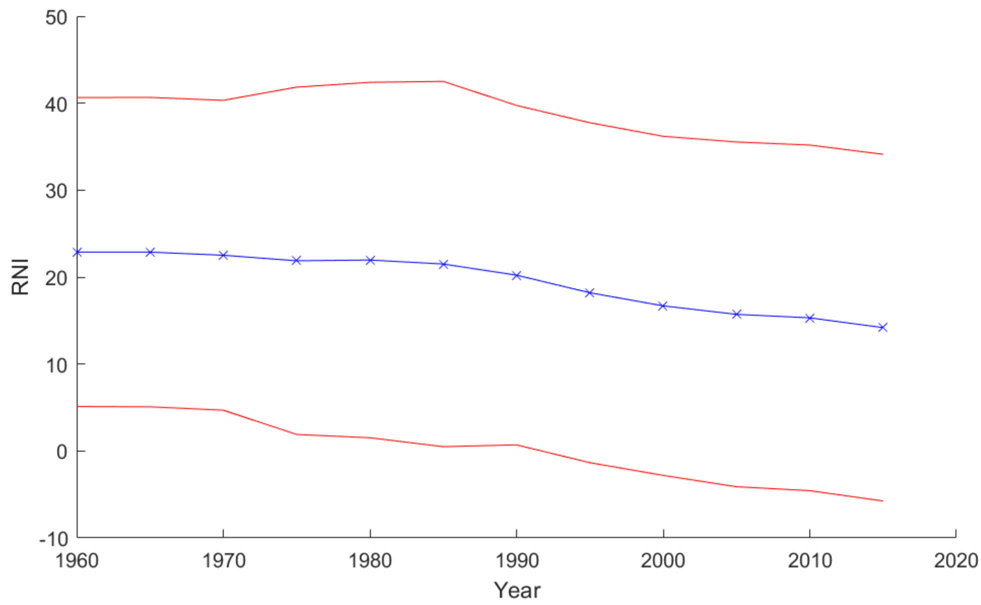

Figure A.2: Rate of Natural Increase (RNI) From 1960 until 2015

Note: The blue line indicates the average rate of natural increase over the period 1960-2015 with 5-year frequency. The red lines indicate the upper and lower bounds of the 95% confidence interval.

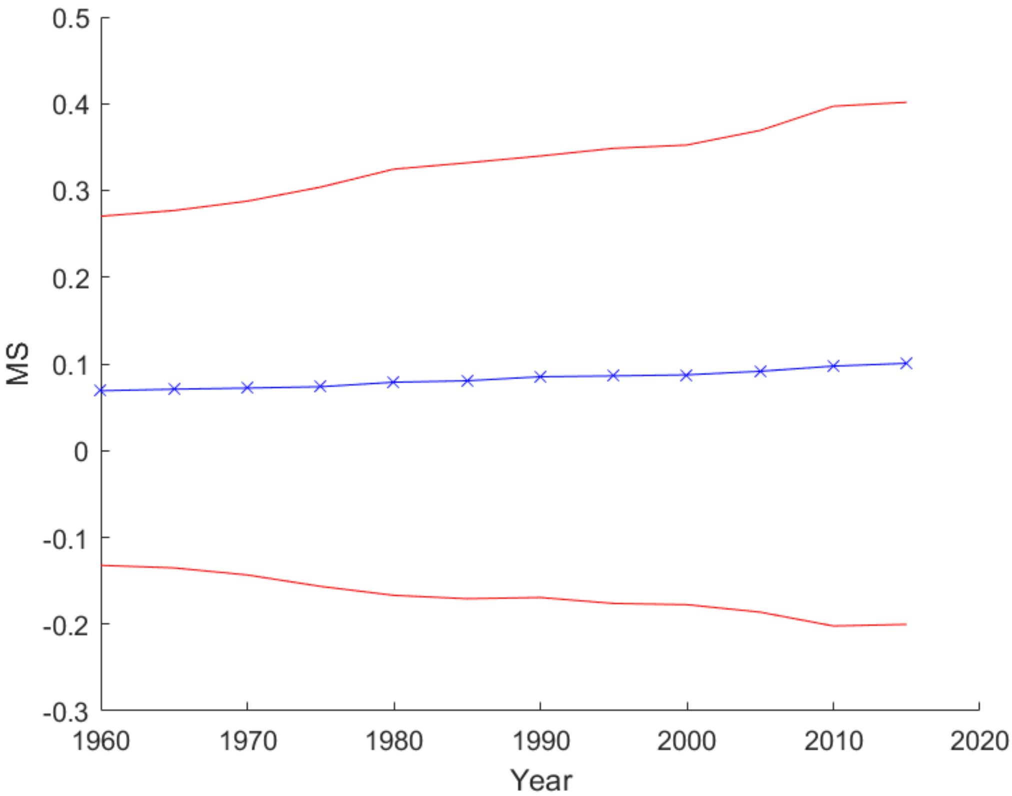

Figure A.3: Average Percentage of International Migrant Stock Relative to the Whole Population From 1960 until 2015

Note: The blue line indicates the average percentage of migrant stock over the period 1960-2015 with 5-year frequency. The red lines indicate the upper and lower bounds of the 95% confidence interval.

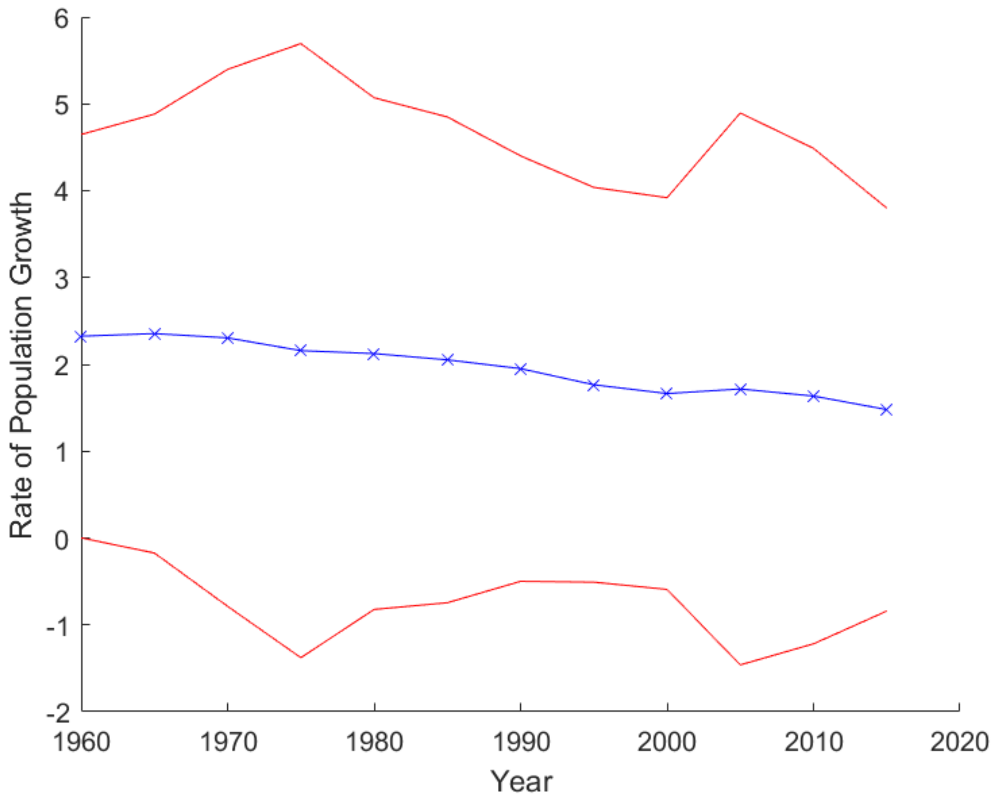

Figure A.4: Average Rate of Population Growth From 1960 until 2015

Note: The blue line indicates the average rate of population growth over the period 1960-2015 with 5-year frequency. The red lines indicate the upper and lower bounds of the 95% confidence interval.

	G				
	1	2	3	4	5
BIC-value	1.1558	1.1788	1.2060	1.2470	1.2640

Table A.1: Corresponding BIC-values for every G=1,...,5

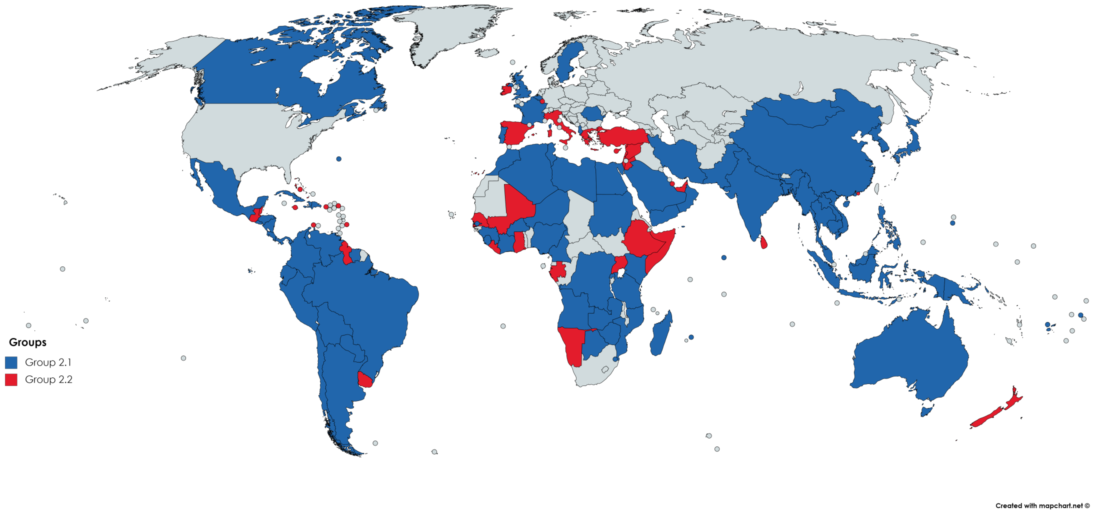

Figure A.5: Division of Countries When Divided into 2 Groups

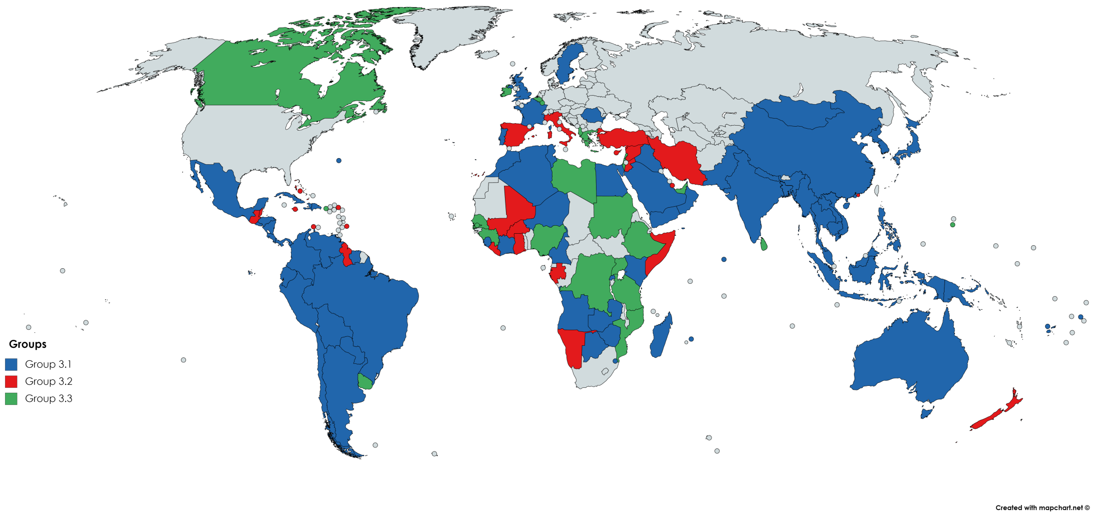

Figure A.6: Division of Countries When Divided into 3 Groups