TI 2020-078/III
Tinbergen Institute Discussion Paper

Dynamic Factor Models with Clustered Loadings: Forecasting Education Flows using Unemployment Data

Francisco Blasques[1,2]
Meindert Heres Hoogerkamp[3]
Siem Jan Koopman[1,2]
Ilka van de Werve[1,2]

[1] Vrije Universiteit Amsterdam
[2] Tinbergen Institute
[3] Dutch Ministry of Education, Culture and Science

Dynamic Factor Models with Clustered Loadings: Forecasting Education Flows using Unemployment Data

Francisco Blasques[1,2,*], Meindert Heres Hoogerkamp[1,3,4], Siem Jan Koopman[1,2] and Ilka van de Werve[1,2,5]

Abstract

We propose a dynamic factor model which we use to analyze the relationship between education participation and national unemployment, as well as to forecast the number of students across the many different types of education. By clustering the factor loadings associated with the dynamic macroeconomic factor, we can measure to what extent the different types of education exhibit similarities in their relationship with macroeconomic cycles. Since unemployment data is available for a longer time period than our detailed education data panel, we propose a two-step estimation procedure. First, we consider a score-driven model which filters the conditional expectation of the unemployment rate. Second, we consider a multivariate regression model for the number of students featuring the dynamic macroeconomic factor as a regressor, and we further apply the k-means method to estimate the clustered loading matrix. In a Monte Carlo study we analyze the performance of the proposed procedure in its ability to accurately capture clusters and preserve or enhance forecasting accuracy. For a high-dimensional, nation-wide data set from The Netherlands, we empirically investigate the impact of the rate of unemployment on choices in education over time. Our analysis confirms that the number of students in part-time education covaries more strongly with unemployment than those in full-time education.

Keywords: Dynamic Factor Models, Cluster Analysis, Forecasting, Education, Unemployment

JEL codes: C38, C53, I25

[1]Vrije Universiteit Amsterdam [2]Tinbergen Institute [3]Dutch Ministry of Education, Culture and Science [4]Dienst Uitvoering Onderwijs (DUO) [5]Netherlands Institute for the Study of Crime and Law Enforcement (NSCR)

[*]F. Blasques is thankful to the Dutch Science Foundation (NWO) for financial support (VI.Vidi.195.099).

1 Introduction

Quality education is one of the Sustainable Development Goals of the United Nations. Emphasizing the importance of education, roughly 11% of total expenditures of the Dutch national budget in 2019 was allocated towards education[1,2]. To secure a reliable budgetary policy, the Dutch government forecasts the numbers of students in each type of education on a nation-wide level. Education systems are complex, dynamic and evolving. To produce accurate forecasts nonetheless, it is important to provide insights into what drives participation in education. In the case of fiscal policy, it is valuable to understand the interaction between education participation and macroeconomic circumstances. Spijkerman (2006) did not find a strong relation between macroeconomic indicators and the total number of students, but macroeconomic circumstances do affect the demand for certain types of education. In particular, the share of part-time education appears to be inversely related with unemployment rates, especially in vocational education. This analysis is relevant because educational institutions receive less funding for part-time students compared to full-time students.

Whether or not distinct groups react differently to macroeconomic circumstances has not been studied in full. More recently, the availability of low-level data allows us to revisit this research question. Our data set for vocational and higher education is high-dimensional on the cross-section but low-dimensional on the time series. The models for such panel data sets typically strike a balance between interpretability and performance. However, to policy makers in government and educational institutions, both interpretability and performance are of interest. A major modeling challenge is to group low-level parameters in a way that we obtain a simpler model that delivers clear and interpretable policy implications while the empirical performance is not compromised, including forecast accuracy.

In this paper, we develop a dynamic factor model where unemployment rates and education participation are modeled simultaneously. Macroeconomic data is typically available for a much longer time period. Hence we can first extract a dynamic economic factor that accounts for the dynamic features in the macroeconomic data set. Since we anticipate that many education flows respond in a similar way to changes in the unemployment rate, we next propose to cluster their dependence on the dynamic economic factor through the parameters of the loading matrix. It

[1] https://www.rijksfinancien.nl/visuele-begroting/2018/owb/u (visual in Dutch), last accessed 2020-04-30.
[2] https://www.rijksoverheid.nl/onderwerpen/financiering-onderwijs/overheidsfinanciering-onderwijs (in Dutch), last accessed 2020-04-30.

imposes a structure on the model that benefits interpretation. Moreover, since we represent many education series by a couple of cluster centroids, it is also more efficient with respect to forecasting education participation.

We propose a two-step estimation procedure for our dynamic factor model. First, we focus on the time series dimension and model the historical unemployment rates through a score-driven local level model as proposed by Creal, Koopman, and Lucas (2013). After estimating the static parameters by maximum likelihood, we extract the dynamic economic factor. This step is important to filter out the noise and preserve the signal in economic data such as the unemployment rate. Second, in the cross-section dimension of our methodology, we take the extracted dynamic economic factor as given. It allows us to model the education data set effectively as a multivariate regression model and estimate the loading parameters by the method of least squares. To gain insights into what types of education respond similarly to changes in unemployment rates, we perform a cluster analysis. By using the k-means method, we are able to represent all loading matrix elements by a few cluster centroids. This ability implies that cluster analysis can support the testing of joint significance for a group of variables without pre-imposition of group compositions. At the same time, we avoid the possible insignificance of individual variables, given that the data for each variable is limited as the time series dimension is small. Once the clusters are identified, we can provide accurate forecasts for all series in the panel.

Dynamic factor models are well-suited to extract common factors from large data sets, see Stock and Watson (2002), Bai and Ng (2002) and Jungbacker and Koopman (2015) amongst others. It has become more prevalent to estimate the parameters in dynamic factor models using a step-wise approach. For example, Doz, Giannone, and Reichlin (2011) first proxy the factors by principal components to estimate the static parameters and then use Kalman filter techniques for signal extraction and forecasting. Bräuning and Koopman (2014) take a slightly different approach, they also first use a principal component analysis for a dimension reduction, but then model all relevant variables jointly in a state space framework such that parameter estimation, signal extraction and forecasting are done by Kalman filter methods. The approaches in both papers are parameter-driven in which the stochastic processes of the factors have their own sources of error.

Our procedure differs in adopting an observation-driven approach: we allow the factors to evolve as dynamic processes which are formulated as functions of past data. In particular, we adopt the approach taken by Creal et al. (2013) and Harvey (2013) where the dynamic specification is based on autoregressive processes with the innovations defined as score functions with respect

to the predictive likelihood function. We first model the unemployment rate data as a score-driven model to filter the dynamic factor, which we then consider as given in the second step of our proposed estimation procedure to estimate the model for the education data. Unrestricted parameter estimates of the loading matrix will be then be clustered to cluster types of education according to their dependence on the unemployment rate. Just as Stock and Watson (2008) do, we use the k-means algorithm in the cluster analysis, although their approach differs in that they base it on the residuals of the dynamic factor model. In our case, we represent the large vector of unrestricted loading estimates by a much smaller vector of cluster centroids. The introduction of clustering within a dynamic factor model has recently also been explored by Hallin and Liška (2011), Barnichon and Mesters (2018), and Alonso, Galeano, and Peña (2020). Their estimation procedures also consist of several steps.

Our simulation study gives a couple of important insights into the performance of the developed dynamic factor model and proposed estimation procedure. First of all, estimation of the full model goes well and the clustered pattern in the loading matrix is correctly identified. Secondly, we find no trade-off with forecast accuracy. We see that clustered forecasting always performances at least as good and can improve the forecast accuracy when there is a clear clustered structure in the data. Even if the the clustered nature is less present, we do not lose much on forecasting accuracy while we gain much on interpretation.

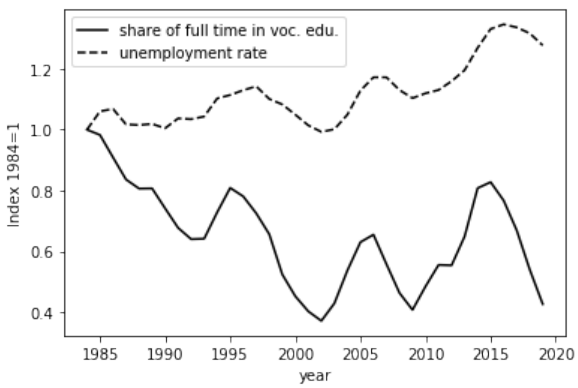
The remainder of this paper is organized as follows. We describe the education matrix and unemployment rate data in Section 2. In Section 3 we propose the dynamic factor model, its two-step estimation procedure and its clustering method for forecasting. Section 4 describes the design and results of our Monte Carlo study. We apply the methodology to our education participation data in Section 5 and show that the proposed methodology can capture important clusters in the flows across the Dutch education system. Section 6 concludes and gives directions for future research.

2 Data

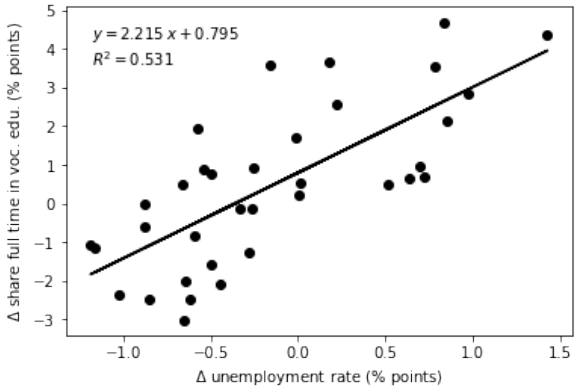
In 2019, 3,7 million students were registered at a Dutch educational institution, around 1.5 million (40.2%) of which in primary education, 960 thousand (25.6%) in secondary education, 500 thousand (13.4%) in vocational education, 460 thousand (12.4%) in higher vocational education (university of applied sciences), and 300 thousand (8.2%) in university. Registry data are aggregated in the

so-called yearly education matrix, that describes flows through the education system. Specifically, it gives the number of people moving from one state (one of 820 education types, 173 diploma type or no education) to another, by age and sex, in a given year. Each state has up to 5 descriptive labels: a sector (e.g. vocational education), type (e.g. on-the-job training, full-time), level, direction (e.g. economics), and grade. Not all flows between the categories are possible (for example, one cannot move from university to primary education) or have not been observed. In total, $33,000$ unique transitions (from origin to destination) have been observed since 2005, and split over age and sex this figure is $326,000$ flows across the education system. Yearly education matrices that are complete for vocational and higher education are available from 2006 onwards. This is a short, wide panel (large N, and small T), implying that many series have to be forecasted while limited data on the time dimension is available.

We enrich the data by including historical macro-level data on the Dutch economy. In particular, we make use of yearly unemployment rates from 1970 to 2019, operationalized as unemployed labor force divided by total labor force; see Bureau for Economic Policy Analysis (2019). In our modeling, we will exploit the fact that the macroeconomic data is available for a much longer period than the education data. Figure 1 provides the instigation for this particular research. Spijkerman (2006) did not find a strong correlation between unemployment and nationwide educational enrollment, but he did show that the growth rate of the share of full-time education covaries positively with the growth rate in unemployment, especially in vocational studies. The left panel indicates that peaks and valleys in the share of full time in vocational studies indeed correspond with the labor market cycle. The right-side panel suggests a linear relationship in first differences.



(a) Share of full-time in vocational education and unemployment rate (indexed at 1984=1)



(b) Increase in unemployment (x-axis) vs. share of full time in vocational education (y-axis)

Figure 1

In this study we are interested in the short-term relation between unemployment rates and inflow into first year of vocational and higher education. We include all inflow from no education and diploma categories. The origin states are grouped by sector, category, and type. Furthermore, to reduce noise, some ages (those that contain fewer first-year students) are binned: $\{< 17, 23-25, 26-30, 31-40, > 40\}$. In 2019, each age category contains a total of between 6,900 and 60,600 students. This leaves $N = 1,155$ time series. In accordance with the scatterplot in Figure 1 the first differences of both datasets are calculated. This also prevents regressing possibly (co)integrated time series. Our macroeconomic time series has length $T_x = 49$, and the education panel has length $T_y = 13$.

3 Methodology

Our education panel data set has many possible education flows which can be selected on the basis of gender and age groups. However, we want to understand the extent to which a macroeconomic variable (the unemployment rate) can help us to forecast the number of students for each category, for a number of years ahead, despite that we only have data available for the last fifteen years. The forecasting method is oftentimes regarded as more convincing when the model preserves a level of interpretability for policy purposes. We therefore develop a dynamic factor model where types of education are clustered based on their dependence on changes in the unemployment rate.

3.1 Modeling framework

Let y_t be the N-dimensional series of education flows in year t. The basic dynamic factor model is given by

$$y_t = \Lambda f_t + \varepsilon_t, \qquad t = 1, \ldots, T_y \tag{1}$$

where Λ is an $N \times k$ loading matrix, $k \times 1$ vector f_t is an unobserved factor and the sequence $\varepsilon_1, \ldots, \varepsilon_{T_y}$ is independent and identically Gaussian distributed with mean zero and variance matrix $\sigma_\varepsilon^2 I_N$. A vector of unit-specific intercepts $\mu = (\mu_1, \ldots, \mu_N)'$ can be added to the model, we then have $y_t = \mu + \Lambda f_t + \varepsilon_t$, but to facilitate the clustered forecasting method in our analysis, we assume that the data is demeaned and we have $\mu = 0$.

Cluster analysis is often used in statistics and machine learning to partition many individuals, cities or products into groups. In our two-step estimation procedure, we propose to cluster the

elements of the loading matrix Λ such that it does not consist of N different elements anymore, but of $K << N$ cluster centroids instead. It does not only imply that we can analyze the similarities and differences between many types of education in their dependence on the unemployment rate, but also that the number of unique forecasts decreases considerably.

To emphasize that the education and macroeconomic data have different time series dimensions we introduce the following notation. The index t and time series length T_y correspond to the education data denoted by y_t, while index $\bar{t}$ and time series length T_x correspond to the unemployment rate data denoted by $x_{\bar{t}}$. The macroeconomic data is typically available for a longer period than the education data, hence $T_x > T_y$, while the time period for the education data coincides typically with the period of the last T_y observations for the macroeconomic data.

We let $x_{\bar{t}}$ be the growth in the unemployment rate in year $\bar{t}$; this time series of yearly observations has length $T_x > T_y$. The location model is given by

$$x_{\bar{t}} = f_{\bar{t}} + \xi_{\bar{t}}, \qquad \bar{t} = 1, \ldots, T_x \tag{2a}$$

where the signal $f_{\bar{t}}$ can be regarded as the time-varying mean of the observed time series $x_{\bar{t}}$, and where $\xi_1, \ldots, \xi_{T_x}$ is assumed to be an independent and identically distributed Gaussian sequence with mean zero and variance σ_ξ^2. This part of our modeling framework enables us to estimate the signal $f_{\bar{t}}$. We use the extracted signal for the estimating of loading parameters in Λ of (1) and for the forecasting of time series variables in y_t of (1). For the filtering of the factor, we adopt the score-driven model as introduced by Creal et al. (2013) and Harvey (2013). In an observation-driven approach, we let

$$\mathcal{X}_{\bar{t}-1} = \{x_1, \ldots, x_{\bar{t}-1}\} = \{\{x_1, \ldots, x_{\bar{t}-1}\}, \{f_1, \ldots, f_{\bar{t}-1}\}\},$$

so that the information set at time $\bar{t}$ is generated by $\{f_{\bar{t}}, \mathcal{X}_{\bar{t}-1}\}$. Since the location model for $x_{\bar{t}}$ in (2a) is linear Gaussian, we have $x_{\bar{t}} \sim p(x_{\bar{t}}|f_{\bar{t}}, \mathcal{X}_{\bar{t}-1}; \theta)$, where $p(\cdot\,|f_{\bar{t}}, \mathcal{X}_{\bar{t}-1}; \theta)$ is the Gaussian density with mean $f_{\bar{t}}$, variance σ_ξ^2, and parameter vector θ. The unknown variance σ_ξ^2 is placed in the parameter vector θ, together with the unknown coefficients that we introduce below. We follow Creal et al. (2013) in their formulation of the filtering or updating equations for $f_{\bar{t}}$; these

are referred to as the generalized autoregessive score (GAS) model and are given by

$$f_{\bar{t}+1} = \omega + \sum_{i=1}^{p} \phi_i f_{\bar{t}+1-i} + \sum_{j=1}^{q} \alpha_j s_{\bar{t}+1-j}, \tag{2b}$$
$$s_{\bar{t}} = S_{\bar{t}} \cdot \nabla_{\bar{t}}, \qquad S_{\bar{t}} = S(\bar{t}, f_{\bar{t}}, \mathcal{X}_{\bar{t}}), \qquad \nabla_{\bar{t}} = \frac{\partial \log p(x_{\bar{t}}|f_{\bar{t}}, \mathcal{X}_{\bar{t}-1}; \theta)}{\partial f_{\bar{t}}},$$

where ω is the intercept, $\phi_1, \ldots, \phi_p$ and $\alpha_1, \ldots, \alpha_q$ are the weight coefficients for the updating mechanism of $f_{\bar{t}+1}$, $s_{\bar{t}}$ is the scaled score with the *local* score function $\nabla_{\bar{t}}$ and the scaling $S_{\bar{t}}$. Typically, we base the scaling on a variance measure of the score function. We refer to this score-driven model by GAS(p, q), where the integers $p \geq 0$ and $q \geq 0$ can be chosen by the econometrician on the basis of fit, residual diagnostics and forecast performance considerations. In many cases, it is sufficient to take $p = q = 1$ and we have the GAS(1,1) model. Given that $p(\cdot|f_{\bar{t}}, \mathcal{X}_{\bar{t}-1}; \theta)$ is the Gaussian density with time-varying mean (or location) $f_{\bar{t}}$, we have

$$\begin{aligned}
\log p(x_{\bar{t}}|f_{\bar{t}}, \mathcal{X}_{\bar{t}-1}; \theta) &= -\tfrac{1}{2}\log 2\pi - \tfrac{1}{2}\log \sigma_\xi^2 - \tfrac{1}{2}(x_{\bar{t}} - f_{\bar{t}})^2/\sigma_\xi^2, \\
\nabla_{\bar{t}} &= (x_{\bar{t}} - f_{\bar{t}})/\sigma_\xi^2, \\
S_{\bar{t}} = \mathcal{I}_{\bar{t}}^{-1} &= \left(-\frac{\partial^2 \log p(x_{\bar{t}}|f_{\bar{t}}, \mathcal{X}_{\bar{t}-1}; \theta)}{\partial f_{\bar{t}}^2} \right)^{-1} = \sigma_\xi^2, \\
s_{\bar{t}} &= \sigma_\xi^2 \cdot (x_{\bar{t}} - f_{\bar{t}})/\sigma_\xi^2 = x_{\bar{t}} - f_{\bar{t}},
\end{aligned}$$

with parameter vector $\theta = \left(\omega\,,\, \phi_1\,, \ldots,\, \phi_p\,,\, \alpha_1\,, \ldots, \alpha_q\,,\, \sigma_\xi^2\right)'$.

Our dynamic factor modeling framework is represented by the equations (1), (2a) and (2b). It facilitates the linkage of the two available data sets (the education panel and the unemployment rate time series) and it provides feasible methods for parameter estimation and forecasting.

3.2 Two-step Estimation Procedure

We propose a two-step estimation procedure for our dynamic factor modeling framework (1), (2a) and (2b). In the first step, we focus on the time series component and use the unemployment rate time series to estimate the score-driven GAS model of equations (2a) and (2b) using the method of maximum likelihood. The static parameters in $\theta = (\omega\,,\, \phi_1, \ldots, \phi_p\,,\, \alpha_1\,, \ldots,\, \alpha_q\,,\, \sigma_\xi^2)'$ are estimated

via the maximization of the loglikelihood function as given by

$$\hat{\theta} = \arg\max_{\theta} T_x^{-1} \sum_{\bar{t}=1}^{T_x} \left(-\tfrac{1}{2}\log 2\pi - \tfrac{1}{2}\log \sigma_\xi^2 - \tfrac{1}{2}(x_{\bar{t}} - f_{\bar{t}})^2/\sigma_\xi^2 \right),$$

where $f_{\bar{1}}$ is initialized by setting it equal to the unconditional mean $\omega/(1-\sum_{i=1}^{p}\phi_i)$, and where $f_{\bar{t}}$ is obtained from the GAS filter (2b), for a given θ and $\bar{t} = 1, \ldots, T_x$. When the maximum likelihood estimate $\widehat{\theta}$ is obtained, we denote the factors obtained from the GAS filter (2b) with $\theta = \widehat{\theta}$ by $\hat{f}_{\bar{t}}$, for $\bar{t} = 1, \ldots, T_x$.

In the second step we consider the education matrix data, focus on the cross-section component, and view equation (1) as a multivariate regression model. We replace the factors f_t by those estimated in the first step, we have $\hat{f}_t \equiv \hat{f}_{\bar{t}}$. The loadings in model equation (1) are estimated by the method of least squares. In this way we obtain an estimate of the variance σ_ε^2 and the unrestricted estimate of the loading matrix as denoted by $\tilde{\Lambda} = (\tilde{\lambda}_1, \ldots, \tilde{\lambda}_N)'$. Next we carry out a cluster analysis on an estimated column of the loading matrix. Using the k-means algorithm we cluster the N different values of a column of $\tilde{\Lambda}$ into K cluster centroids for $\hat{\Lambda}$ as follows:

1. Initialize the cluster centroids randomly from the data range: draw centroids $\delta_1, \ldots, \delta_K$ from $U(\tilde{\Lambda}_{min}, \tilde{\Lambda}_{max})$, where $U(\cdot)$ is the uniform distribution
2. Obtain distances between data points and cluster centroids and add cluster labels to the data points: label $c^{(i)} = \arg\min_k ||\tilde{\lambda}_i - \delta_k||^2, \forall i$
3. For each cluster, assign new centroids: centroid $\delta_k = \frac{\sum_{i=1}^{N} \mathbb{1}\{c^{(i)}=k\}\tilde{\lambda}_i}{\sum_{i=1}^{N} \mathbb{1}\{c^{(i)}=k\}}, \forall k$
4. Verify whether the cluster centroids changed
5. Repeat steps 2-4 until convergence
6. Replicate steps 1-5 for $R = 15$ different random seeds
7. Keep clustered loading matrix $\hat{\Lambda}$ with shortest total distance to the cluster centroids

The consistency of the maximum likelihood estimates of the parameters in the score-driven model, in the first step, can be obtained by application of standard theory as delivered by Blasques, Koopman, and Lucas (2016). The consistency of the estimates of the clustered loadings, in the second step, is corroborated by the Monte Carlo study developed in this paper, and can be obtained using asymptotic results for plug-in M-estimators as detailed in Chen and Liao (2014); we leave this asymptotic-theoretical characterization of the second-step estimator for future research.

3.3 Clustered Forecasting

After the two-step estimation procedure, we use the estimated static parameters and filtered time-varying factors of the score-driven model to forecast future values of the factors in a recursive manner from equation (2b). Next, together with the estimated clustered loading matrix, we forecast future education participation. Since the clustered loading matrix $\hat{\Lambda}$ consists of K distinct values only, we just need to forecast $K << N$ series. This leads to computation time savings because we have thousands of education flows for each combination of age and gender.

4 Monte Carlo study

We carry out a Monte Carlo study to verify the performance of our estimation and measurement methodology. For this study, we consider the dynamic factor model as given by the equations equations (1) to (2b) with the GAS updating orders $p = 1$ and $q = 1$, that is the GAS(1,1) specification. The dimensions in our simulation design are motivated by the empirical problem at hand. Hence we set the cross-section dimension of y_t to be relatively large and the time series dimension to be relatively small. In particular, we have $N = 500$ and $T_y \in \{10, 20\}$. We set the time series dimension of $x_{\bar{t}}$ to be moderate, with $T_x \in \{50, 100\} > T_y$. To verify the forecasting performance, we set the forecast horizon to be $F = 3$ periods ahead. Moreover, we take the static parameters in the score-driven model as $\sigma_\xi = 1, \omega = 0.3, \phi = 0.95$ and $\alpha = 0.1$, making the unconditional mean of the process for the stationary factors equal to $\omega/(1-\phi) = 6$. The five equally-sized clusters of the loading matrix have centroids 4, 11, 19, 23 and 35. This implies that the observations roughly vary between $4 \times 6 = 24$ and $35 \times 6 = 210$. Finally, we set the error variance to follow from $\sigma_\varepsilon = 20$. The reported Monte Carlo results in our study are based on $M = 1,000$ simulations, for the different model specifications and data dimensions.

The specific choices of the static parameters and cluster centroids are for illustrative purposes. We do need to assume that there is some underlying form of clustering in the data present and we can visualize that by taking a non-zero unconditional mean of the factors and distinct choices of the centroids. It is also a realistic assumption in our empirical study; although the number of possible education flows is huge, it is likely that many behave alike as their characteristics are alike. However, the behavior over time can vary between the series within a cluster. To visualize that, we present a couple of example time series with varying variance σ_ε^2 in Figure 2 for $T_x = 100, T_y = 10$, two clusters with loadings 23 and 35 and remaining settings as above.

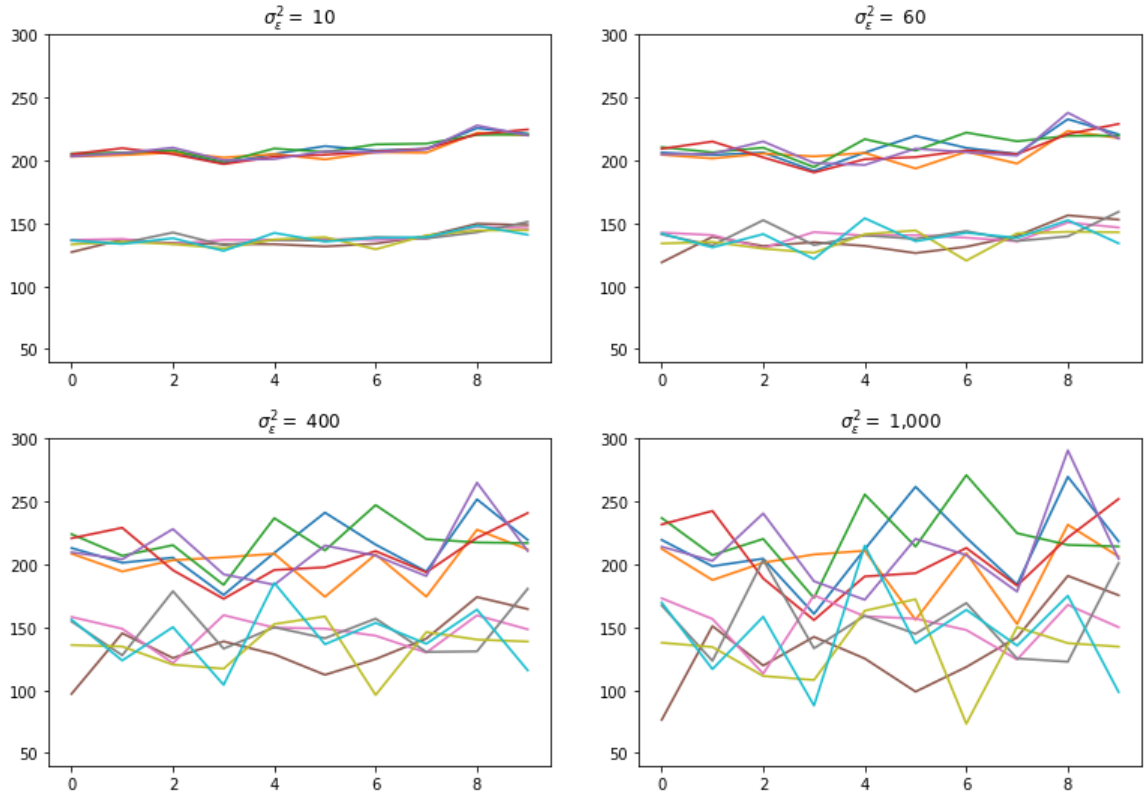


Figure 2: Example time series y_t plotted against time. The same five time series from two clusters with loadings 23 and 35 are given in all plots. Only the variance σ_ε^2 differs over the plots, the other simulation settings are kept fixed ($T_x = 100, T_y = 10, N = 500, \sigma_\xi = 1, \omega = 0.3, \phi = 0.95, \alpha = 0.1$).

In these plots we see the trade-off when clustering is beneficial and when not. For a very small variance, such as the $\sigma_\varepsilon^2 = 10$ in the top-left plot, there is no real need for imposing the cluster analysis. By basically scanning over such data plots, one gains already the knowledge on patterns in the data. But also practically, as there is basically no variation over time, the unrestricted estimate will be as good as the clustered one. In the other extreme, where the variance is very large as in the bottom-right plot with $\sigma_\varepsilon^2 = 1,000$, clustering is also not useful because there are no clusters to really distinguish. The large variation over time will make getting an unrestricted estimate already challenging and the cluster classification is therefore also difficult. More reasonable values in-between, such as the plots with $\sigma_\varepsilon^2 \in \{60, 400\}$, show where extracting the clustered pattern can be of added value. With some variation over time, the unrestricted estimates may be a bit too far off in either direction for each of the series, however, this is offset by using clustered estimates. In such cases, one might think that the unrestricted estimates are very different, but the clustered estimate clearly shows that their behavior is actually similar. In the discussion of the full simulation study next, we will continue with $\sigma_\varepsilon^2 = 400$.

4.1 Parameter estimation results

The performance of the two-step estimation procedure will be visualized by densities of the estimated parameters and cluster centroids. We will judge the clustering classification by confusion matrices[3]. Furthermore, we give in-sample statistics to compare the fit of the unrestricted and clustered models. To judge the accuracy of clustered forecasting, we compute loss functions of forecasting with and without clustering. We then consider fraction of the two counts and prefer clustered forecasting if

$$AF = M^{-1} \sum_{m=1}^{M} \left(\frac{LF_m^{clustered}}{LF_m^{unrestricted}} \right) < 1,$$

where the numerator reflects clustered forecasting (with $\hat{\Lambda}$ after running k-means) and the denominator reflects unrestricted forecasting (with $\tilde{\Lambda}$ directly after least squares) for loss function $LF \in \{MSE, MAE\}$.

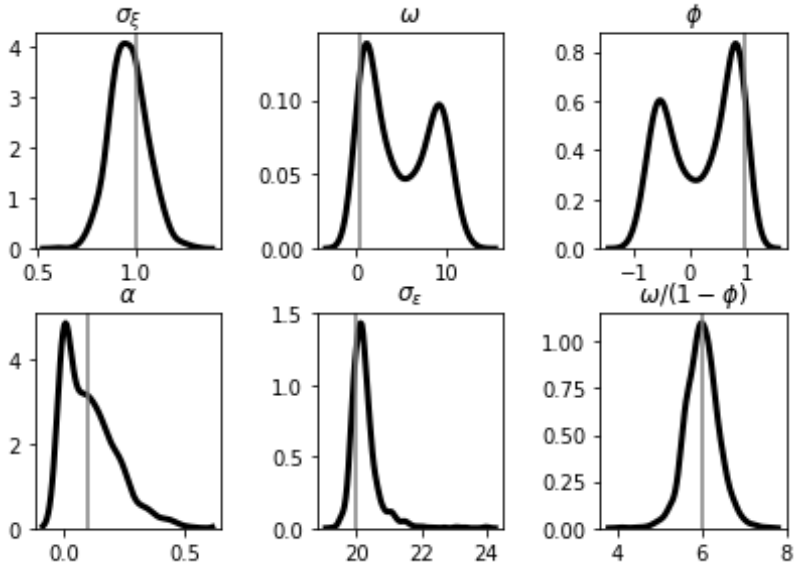
Figures 3 and 4 give the density plots of the estimated static parameters and cluster centroids. For both figures we take $T_y = 10$ fixed and first consider $T_x = 50$ and then $T_x = 100$. Similar results for $T_y = 20$ are given in the appendix. In each figure, the plotted densities of the estimated static parameters are given in the first set of results and the plotted densities of the cluster centroids in the loading matrix in the second set of results, all based on $M = 1,000$ simulations.

If we focus on the first step of our proposed estimation procedure, then we consider the static parameters σ_ξ, ω, ϕ and α. Only the time series dimension T_x is of importance for these parameters. Comparing the corresponding densities in Figures 3 and 4 (or similarly comparing Figures A.1 and A.2 because the varying time series dimension T_y is not relevant in this step) clearly shows that the parameter estimates become much more precise as the time series dimension T_x increases. This would improve further if we take T_x even larger, but that does not match our empirical study, so we leave that out here.

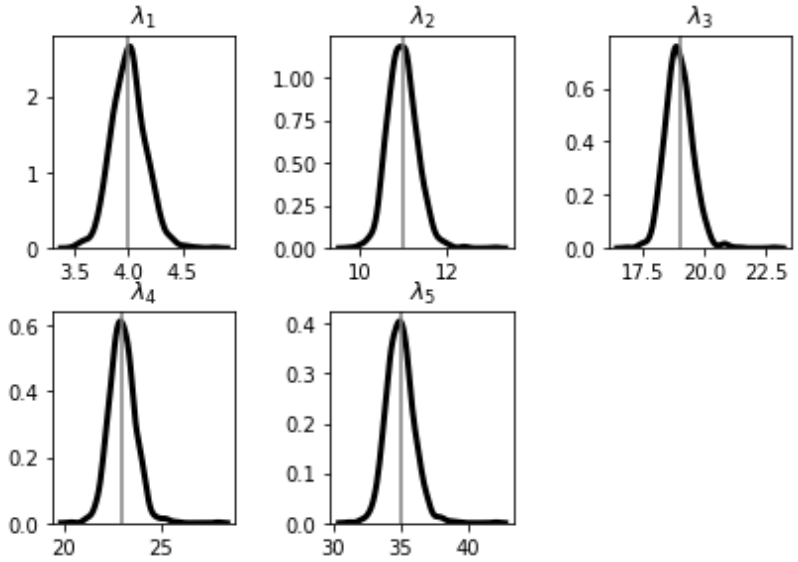
What is mostly apparent in the density plots for $T_x = 50$, but also a bit for $T_x = 100$, is that the densities of ω and ϕ seem bi-modal. Since the density plot of the unconditional mean $\omega/(1-\phi)$ is uni-modal around the true value, it might suggest that an identification issue is present in the first step for smaller time series dimensions. But since these apparent biases offset each other, the filtered factors are still well estimated and it does not have any further consequences for our parameter of interest, that is, the cluster centroids. All estimated cluster centroids are centered

[3] A confusion matrix gives insight on the correct assignment to the clusters instead of the specific values of the centroids. It gives the counts of correct and incorrect assignments to the clusters with smallest, second-to-smallest, ..., largest centroid. Perfect classification would give a diagonal matrix.

around their true values with just small deviations.

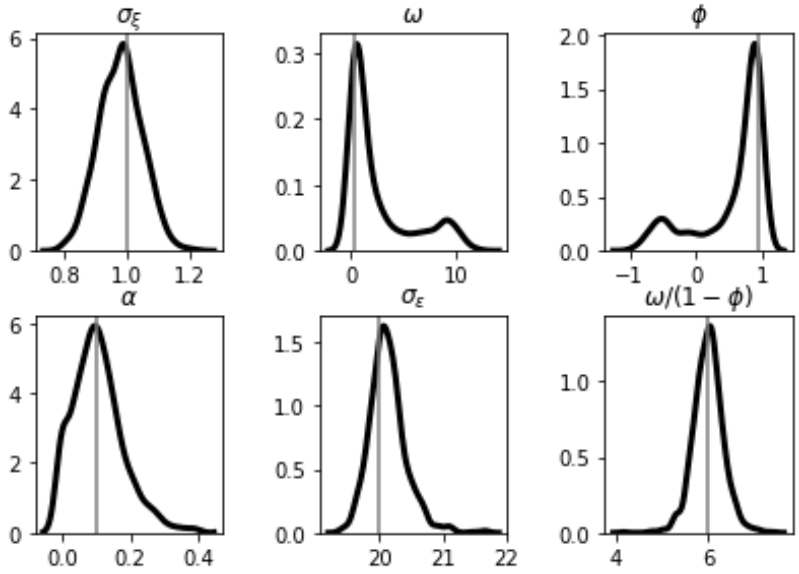


(a) Densities of estimated static parameters.

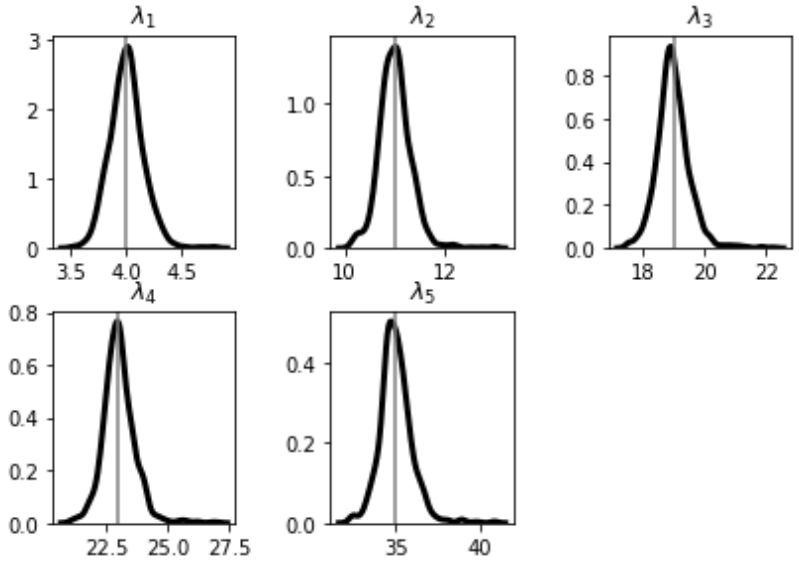


(b) Densities of estimated cluster centroids.

Figure 3: Parameter estimation results of $M = 1,000$ simulations for $N = 500, T_y = 10, T_x = 50$. Vertical lines represent true values.



(a) Densities of estimated static parameters.



(b) Densities of estimated cluster centroids.

Figure 4: Parameter estimation results of $M = 1,000$ simulations for $N = 500, T_y = 10, T_x = 100$. Vertical lines represent true values.

Besides the estimated values of the cluster centroids, the assignment to the correct cluster is also of importance. For that purpose we present confusion matrices in Table 1. In the columns of this Table we vary $T_y \in \{10, 20\}$, while in the rows $T_x \in \{50, 100\}$ varies. First we recall that the results of the first step of our proposed estimation procedure only depended on the time series dimension T_x. So, the time series dimension of the second step here, denoted by T_y, is now of interest. Already for $T_y = 10$ the confusion matrices look good, with lower boundaries of 96.9%,

and for $T_y = 20$ the lower boundaries are even 99.6%. This confirms our earlier finding that even though there might be an identification issue in the first step and the corresponding parameters might be somewhat biased, it does not lead to any problem in identifying the clusters and cluster centroids in the second step, meaning that we can correctly identify the structure in the data.

	$\mathbf{T_y = 10, T_x = 50}$					$\mathbf{T_y = 20, T_x = 50}$				
	E1	**E2**	**E3**	**E4**	**E5**	**E1**	**E2**	**E3**	**E4**	**E5**
C1	99.960	0.040	0	0	0	100	0	0	0	0
C2	0.057	99.931	0.012	0	0	0.001	99.999	0	0	0
C3	0	0.009	96.938	3.053	0	0	0	99.574	0.426	0
C4	0	0	2.954	97.046	0	0	0	0.378	99.622	0
C5	0	0	0	0	100	0	0	0	0	100

	$\mathbf{T_y = 10, T_x = 100}$					$\mathbf{T_y = 20, T_x = 100}$				
	E1	**E2**	**E3**	**E4**	**E5**	**E1**	**E2**	**E3**	**E4**	**E5**
C1	99.952	0.048	0	0	0	100	0	0	0	0
C2	0.043	99.945	0.012	0	0	0	100	0	0	0
C3	0	0.010	96.925	3.065	0	0	0	99.602	0.398	0
C4	0	0	2.996	97.004	0	0	0	0.383	99.617	0
C5	0	0	0	0	100	0	0	0	0	100

Table 1: Confusion matrices of estimated cluster centroids. All results are based on $M = 1,000$ simulations with $N = 500$. The top panels have $T_x = 50$ while $T_y \in \{10, 20\}$ varies and the bottom panels have $T_x = 100$ fixed while $T_y \in \{10, 20\}$ varies. In each panel, the row labels indicate the true clusters and the columns labels to the assigned clusters in estimation. Frequencies are given (here also equal to percentages), perfect classification would be $100I_5$. From smallest to largest, the true values are $4, 11, 19, 23$ and 35.

The estimated cluster centroids and confusion matrix are rather precise. However, our key interest is the comparison between the unrestricted model and the clustered model. For that matter, we present in-sample statistics in Table 2 for $T_y = 10$. Similar results for $T_y = 20$ are given in the appendix. In the simulation study, we used five equally-sized clusters with centroids $4, 11, 19, 23$ and 35 as loading matrix. This Table reports for each of the clusters, and the full sample, the mean squared error and mean absolute error of the unrestricted estimates and estimated cluster centroids compared to the true values and the difference between the two estimates. Overall, the estimated cluster centroids always outperform the unrestricted estimates. Furthermore, the Table also reports the contribution of each cluster to the log likelihood. We can then compare the models by the AIC and, since the clustered model is a restricted version of the unrestricted model, the LR-test. In all cases, the LR-test provides evidence that there is no significant difference between the log likelihood values. However, taking into account that the clustered model has much less parameters than the unrestricted model, the AIC shows that the clustered model should be preferred. In both Tables,

the differences between the clusters are small because the clusters are here chosen to be the same in size and without deviation around the cluster centroid, but in empirical studies such statistics will give insight on the homogeneity of the different clusters.

$\mathbf{T_y = 10, T_x = 50}$

	MSE			MAE			Unrestricted		Clustered		
	Unr.	**Cl.**	**Diff.**	**Unr.**	**Cl.**	**Diff.**	**LL**	**AIC**	**LL**	**AIC**	**LR**
C1	1.134	0.047	1.106	0.849	0.127	0.839	-4,324	8,850	-4,375	8,754	102
C2	1.234	0.150	1.117	0.884	0.267	0.842	-4,325	8,852	-4,376	8,756	102
C3	1.432	0.827	1.015	0.948	0.538	0.817	-4,332	8,866	-4,378	8,760	92
C4	1.575	0.929	1.008	0.992	0.623	0.813	-4,335	8,872	-4,381	8,766	92
C5	2.168	1.053	1.115	1.157	0.786	0.839	-4,350	8,902	-4,399	8,802	98
Full	1.508	0.601	1.072	0.966	0.468	0.830	-21,884	44,770	-22,130	44,272	492

$\mathbf{T_y = 10, T_x = 100}$

	MSE			MAE			Unrestricted		Clustered		
	Unr.	**Cl.**	**Diff.**	**Unr.**	**Cl.**	**Diff.**	**LL**	**AIC**	**LL**	**AIC**	**LR**
C1	1.137	0.047	1.111	0.849	0.121	0.840	-4,323	8,848	-4,374	8,752	102
C2	1.216	0.134	1.109	0.876	0.243	0.839	-4,323	8,848	-4,374	8,752	102
C3	1.408	0.802	1.010	0.937	0.508	0.814	-4,326	8,854	-4,373	8,750	94
C4	1.519	0.888	1.004	0.971	0.575	0.811	-4,328	8,858	-4,375	8,754	94
C5	2.054	0.950	1.104	1.114	0.705	0.837	-4,338	8,878	-4,388	8,780	100
Full	1.467	0.564	1.067	0.949	0.430	0.828	-21,856	44,714	-22,105	44,222	498

Table 2: Model fit for unrestricted and clustered model. All results are based on $M = 1,000$ simulations with $N = 500$ and $T_y = 10$, with $T_x = 50$ in the top panel and $T_x = 100$ in the bottom panel. Each row in a panel represents a cluster and the last row is the full sample. In the columns are loss functions given of the unrestricted loadings compared to the true ones ("Unr."), the clustered centroids compared to the true ones ("Cl.") and the unrestricted loadings minus the clustered centroids ("Diff."). The first three columns have the MSE as loss function and the last three columns the MAE. For both models, the log likelihood and AIC are given and they are compared via the LR-statistic in the last column. For the latter, the critical values are 123 for each cluster (100-1 degrees of freedom) and 548 overall (500-5 degrees of freedom), at the 5% significance level.

4.2 Forecasting results

By enforcing the clustered structure in our modeling framework, we do not want to compromise on the forecasting performance. We recall that our proposed procedure is preferred if the average fraction AF in Section 4.1 is smaller than unity. Table 3 reports these average fractions; in the columns we vary $T_y \in \{10, 20\}$, while in the rows we vary $T_x \in \{50, 100\}$. Furthermore, the row f in Table 3 reports the performance for forecasting $f \in \{1, 2, 3\}$ steps ahead.

The time series dimension T_x of the first step in our proposed estimation procedure does not have a big impact on the estimation results of the second step, this is also confirmed by the forecasting

results. An improvement of more than 4% in the mean squared error is obtained if $T_y = 20$. This becomes 7% if T_y is only half of it. This reveals the strength of our proposed procedure: for data sets where forecasting is of interest but the time series dimension is small although the cross-section dimension is large, it is beneficial to exploit the clustered nature of the data. As the time series dimension T_y increases, clustered forecasting goes to unrestricted forecasting because the unrestricted estimates of the loadings become less biased in the second step. While for short time series dimensions T_y, the clustering averages out these biases such that the forecasting performance improves.

		$\mathbf{T_y = 10}$		$\mathbf{T_y = 20}$	
		MSE-ratio	*MAE-ratio*	*MSE-ratio*	*MAE-ratio*
$\mathbf{T_x = 50}$	*f = 1*	0.930	0.963	0.957	0.978
	f = 2	0.929	0.963	0.957	0.978
	f = 3	0.929	0.963	0.958	0.978
$\mathbf{T_x = 100}$	*f = 1*	0.928	0.962	0.956	0.978
	f = 2	0.928	0.962	0.956	0.977
	f = 3	0.930	0.963	0.958	0.978

Table 3: Forecasting performance of clustered forecasting versus unrestricted forecasting. All results are based on $M = 1,000$ simulations with $N = 500$. The top panel has $T_x = 50$ while $T_y \in \{10, 20\}$ varies and the bottom panel has $T_x = 100$ fixed while $T_y \in \{10, 20\}$ varies. Clustered forecasting is preferred if $M^{-1} \sum_{m=1}^{M} \left(\frac{LF^{clustered}}{LF^{unrestricted}} \right)_m < 1$, where the numerator reflects clustered forecasting (with $\hat{\Lambda}$ after running k-means) and the denominator reflects unrestricted forecasting (with $\tilde{\Lambda}$ directly after least squares) for loss function $LF \in \{MSE, MAE\}$. The first row in each cell represents one-step ahead forecasting, the second row two steps ahead and the last row three steps ahead.

In the simulation study, the time series dimensions T_x and T_y and the cross-section dimension N were chosen in line with the empirical study, however, the static parameters and loadings were chosen under the assumption that the data consists of clearly identifiable clusters. In practice, this might not be the case. For example, in our empirical study, we have that most education flows are spread around zero and there are some larger values. In that case, the unrestricted estimates of the loadings and the cluster centroids will be close to each other and hence our performance measure of forecasting will go to one. This would mean that unrestricted and clustered forecasting should be equally preferred if one is only interesting in the forecasts themselves. However, in the policy relevant context of our empirical study, we are especially interested in the interpretation. By the structure that we put on the model, we gain a lot on interpretation while we don't pay for forecasting performance. This in combination with the decent improvements in forecasting performance that we found in the simulation study above gives enough evidence that our methodology can also be generalized and applied in different contexts in future research.

5 Empirical study

In a study on education enrollment in the Netherlands, Spijkerman (2006) found that educational choices are related to unemployment rates, in particular in part-time and on-the-job education. In the literature, proposed causal relations usually concern demand for education vis-a-vis supply of labor. Economists typically assume substitution: people allocate their time towards education and the labor market. In this line of thought, enrollment is understood as an investment decision (Clark, 2011, pp. 524-525). Labor market characteristics such as vacancies, unemployment rates, and wages, influence the choice made at any given moment. For example, higher demand for labor increases the opportunity costs of education, decreasing its relative preference. Conversely, when confronted with a weak labor market, young students are more likely to remain in education; see i.a. Lamb, Walstab, Teese, Vickers, and Rumberger (2004), Clark (2011, p. 523). For post-initial education, a tight labor market might induce to re-educate, looking for better job prospects or to protect their position (Groenez, Desmedt, & Nicaise, 2007). There are effects on the supply-side of education as well, especially in on-the-job learning. When demand for labor is high, employers are more likely to provide apprenticeships opportunities. This relationship is assumed to have multiple causes: apprenticeships can be a substitute for hard to find workers (especially for middle-skill vacancies), or they are a way to attract talent. This is the reasoning behind the correction for unemployment vocational education in the Dutch student forecasts (Ministry of Education, Culture and Science, 2020).

A drawback of some of this research is that it does not recognize non-stationarity in the regressed time series (i.a. participation in education and economic indicators); see i.a. Lamb et al. (2004, pp. 126-132). As a result, regressions might be spurious. Also, many suggested causal relations are based on cross-country comparative analysis (e.g. OECD, EU), in which participation choices of individuals cannot be distinguished from the varying institutional landscapes (Groenez et al., 2007, p. 2). In this research, we use data from the Netherlands only, and increase N by lowering the level of analysis; i.e. by studying transitions from one type of education to another. Specifically, the change in unemployment rate is regressed on the changes in transitions into first grade studies in Dutch vocational and higher education. Figure 1 suggests a linear relation in differenced unemployment rates and differenced share of full-time students in vocational education. In the following analysis, we will not test causal claims, for which a structural causal model should be developed.

Transitions with a higher number of students on average have a higher variance. To normalize,

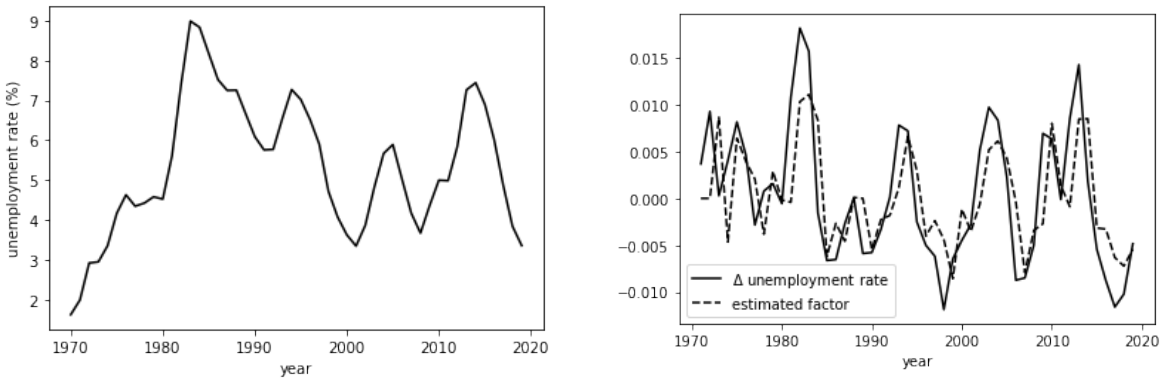


Figure 5: Unemployment rate and estimated factor. Left: unemployment rate at $T = 1970, ..., 2019$. Right: differenced unemployment rate and estimated factor $\hat{f}_t$ at $T = 1971, ..., 2019$

	GAS(1,1) w/o ω	GAS(1,1)	GAS(2,1) w/o ω	GAS(1,2) w/o ω
$\hat{\sigma}^2_\xi$	0.005	0.005	0.005	0.005
$\hat{\omega}$	-	-0.000	-	-
$\hat{\phi}_1$	0.365	0.365	0.355	0.329
$\hat{\phi}_2$	-	-	-0.117	-
$\hat{\alpha}_1$	0.933	0.937	0.956	0.936
$\hat{\alpha}_2$	-	-	-	0.039
logL	-181.90	-181.06	-181.35	-181.01
AIC	369.80	370.12	370.70	370.02

Table 4: Parameter estimates and metrics for several GAS specifications.

each cross-sectional unit is divided by its average level: $\tilde{y}_{it} = \Delta y_{it}/\bar{y}_i$, with y_{it} the number of people in transition i at time t. Several specifications of the GAS model have been tested; see Table 4. Based on the AIC, for our yearly observed macroeconomic time series, the GAS(1,1) without intercept seems to be most suitable. The filter is initialized using the unconditional mean (0). The right panel in Figure 5 presents the differenced unemployment rate (solid) and the estimated common factor (dashed) over time.

With $\hat{\phi}_1 = 0.365$, extrapolating from this filter results in rapidly regression towards the mean (0). Comparing clustered and unclustered loadings will not be very meaningful when the forecasted factor is close to 0. Moreover, the model with one factor is sensitive to variance in the forecast of x_t. Because we do not have much leeway to vary T and F (short panel), the out-of-sample tests are thus less reliable[4]. Instead, we will rely on in-sample metrics to compare the clustered and

[4]In an exploration of forecasting performance on random subsets of the data, we found the clustered model did not underperform for the unrestricted one. MSE-ratios center closely around 1.0. More research is needed to draw conclusions.

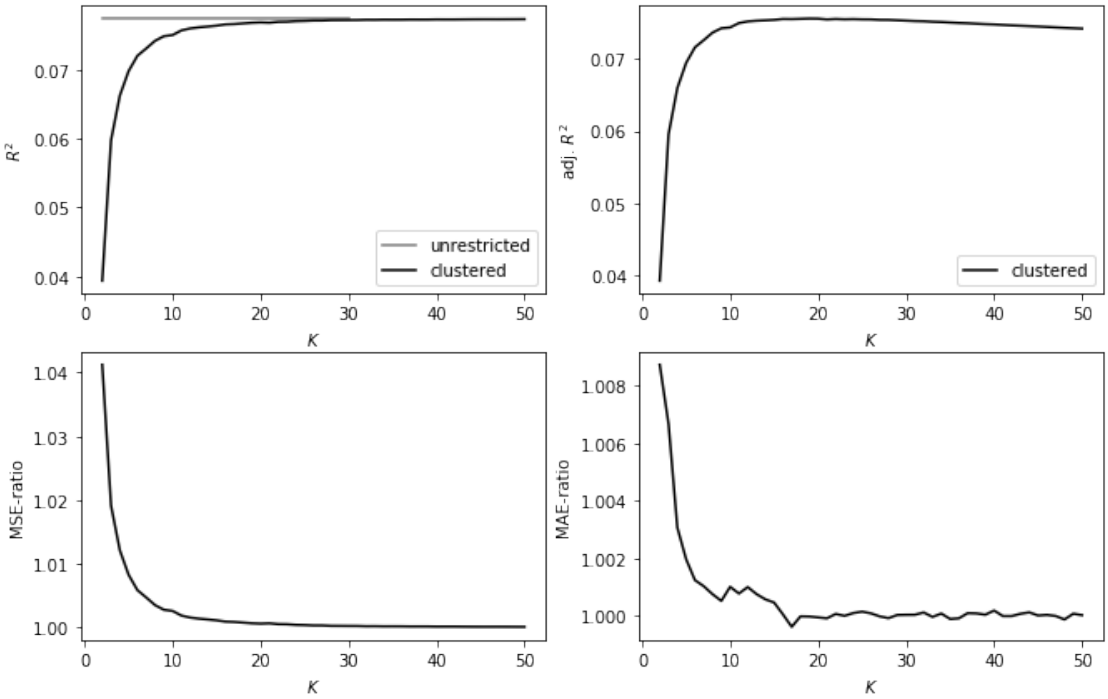


Figure 6: Various metrics (R^2, $\bar{R}^2$, MSE-ratio and MAE-ratio) for $K = 2, ..., 50$ (with $R = 1{,}000$ repetitions using different centroid seeds)

unrestricted models.

The unrestricted model reveals that the unemployment rate explains about 7.7% of the variation in the education flows in y_t. This low number is to be expected. Indeed, we expect that unemployment is a relevant factor for only a part of the transitions into education. The R^2 of the restricted model is naturally lower, but it converges to this number as the number of clusters increases. Likewise, MSE- and MAE-ratios in the right panel converge to 1, see Figure 6. Increasing the number of centroids has diminishing returns in predictive performance. To find the optimal number of clusters, the adjusted R^2 ($\bar{R}^2$) is used, reaching its maximum is at $K = 19$. The MSE-ratio is 1.001, meaning that the predictive performance of the clustered model is almost on par with that of the unrestricted one. Thus, clustering the loading matrix is an effective way to reduce the number of parameters in the model. Other ways to select the optimal number of clusters include maximizing the AIC or the 'elbow method' (see section 3.2).

Figure 7 provides a representation of the estimated loading coefficients. It depicts the cumulative distribution of unrestricted loadings and a step-wise cumulative distribution of clustered loadings. The estimated unrestricted have sample average -2.67 (std.dev. 13.95), the clustered loadings have average -2.68 (std. dev. 13.89). The smallest and largest loading are -60.58 and 47.32 respectively,

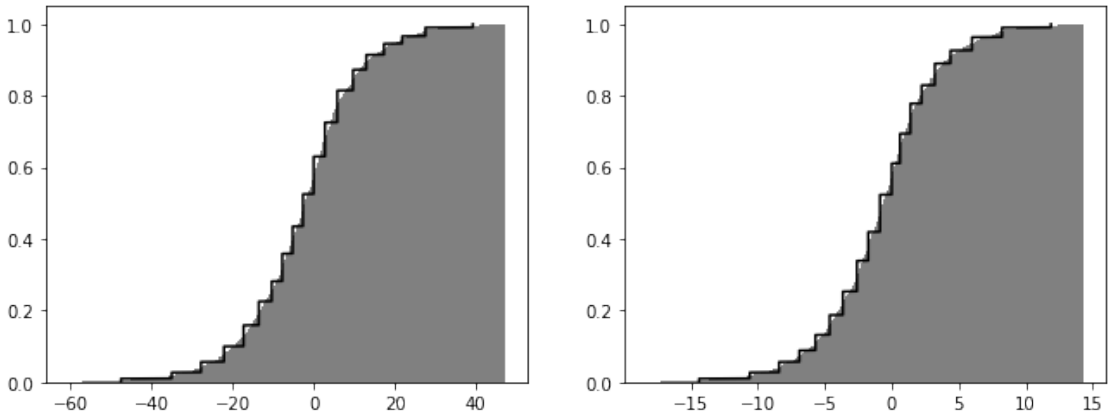


Figure 7: Cumulative density of loadings. On the left with OLS, on the right with ridge regression ($\alpha = 0.01$). The stepwise line indicates the position of clusters and distribution of clustered loadings.

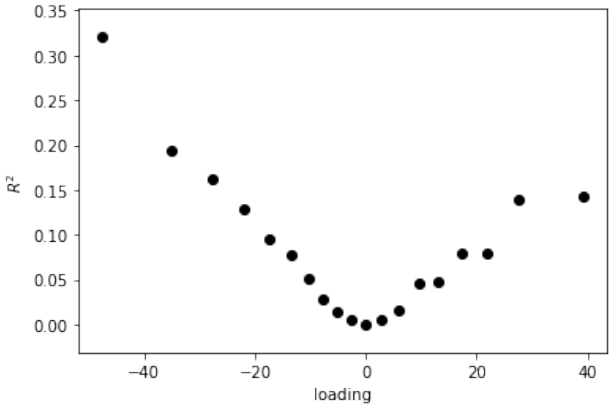


Figure 8: R^2 per cluster

meaning that for these transitions 1% change in the unemployment factor corresponds to a ¿40% change (relative to a transition's sample average) in student counts. No obvious clusters emerge from the unrestricted loadings, which are unlikely to exist in the real world. Still, clustering might be helpful in situations where model simplicity is attractive, which is often the case in public policy analysis. The main advantage of this model is that decreasing the number of estimated coefficients reduces model complexity, without losing notable explanatory performance.

Figure 8 presents the R^2 per cluster. As one would expect, the common factor is most relevant for clusters with larger loadings. In one cluster, 32.1% of variance is explained by the filtered unemployment factor. In the left tail we find transitions into vocational education, which seem to be negatively related to unemployment rates. This is in support of the hypotheses that participation in on-the-job learning depends on the availability of apprenticeships position and/or that low unemployment induces people to learn market-oriented skills. Similarly, the right tail of the

distribution contains many transitions into school-based vocational education, moving pro-cyclical with unemployment.

Across the panel, we find that part-time/on-the-job education tends to covary negatively with unemployment; see Table A.2. Similarly, the results indicate that unemployment rates affects people aged 31 and above most; see Figure 9. Also, those not in education are less likely to study with increasing unemployment, whereas transitions from diploma origins do not show a strong covariance. Thus, the results do not suggest that a weak labor market induces people to reorient. Instead, the results favor the notion that people are more likely to participate in post-initial education when having better job prospects. Alternatively, post-initial education depends on the availability of apprenticeship positions. Although clustering strongly reduces the number of parameters, Figure 9 indicates that the structure of the loading matrix stays intact. Loadings also vary to some extent across directions in education, but these results do not provide enough basis to make conclusions.

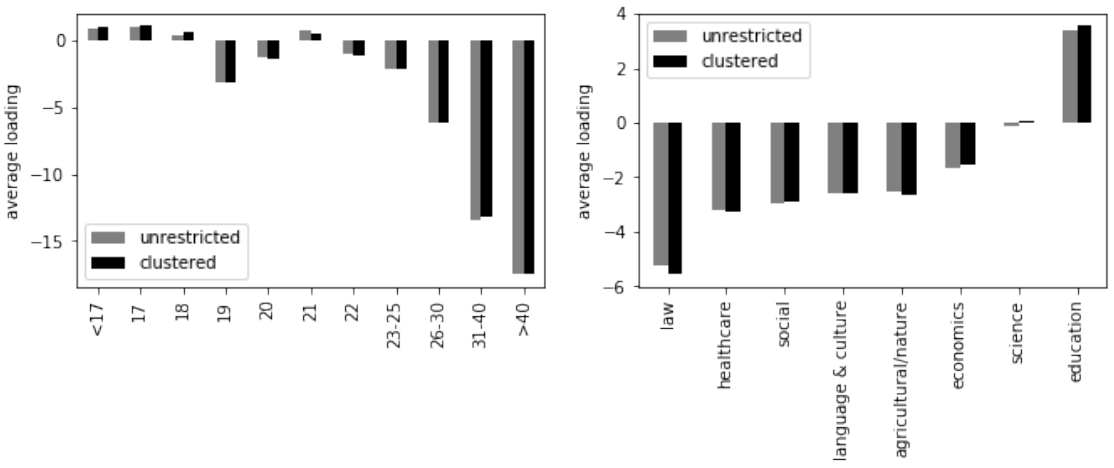


Figure 9: Average loadings in unrestricted and clustered model (weighted by average level of the transition $\bar{y}$) for age (groups) and directions

A drawback of the linear regression model on the short panel is that estimated parameter coefficients are largely affected by random noise. One way to reduce model variance is to include a penalty in the objective criterion. For an illustration, see the right panel of Figure 7 for the distribution of estimated loadings using ridge regression with $\alpha = 0.001$). Model variance might also be reduced by using the descriptive labels of the educational time series. The categories could be included in a k-means type algorithm; see e.g. Huang (1998) for an extension to cluster in a setting with mixed categorical and numerical data.

We end this analysis by plotting impulse response functions (IRFs) produced by a unit shock

in the differenced unemployment rate x_t at time $t = 0$ ($\xi_0 = 1$). Figure 10 plots these IRFs which show the dynamic impact of the unemployment shock in education flows. Through the updating equation, the common factor f_t responds with 1 step delay. Additionally, the education flows predicted by the linear regression model covary synchronously with f_t. Some of the clusters covary positively and some negatively. The right panel shows the effects on the differenced transitions into education for the 19 clusters.

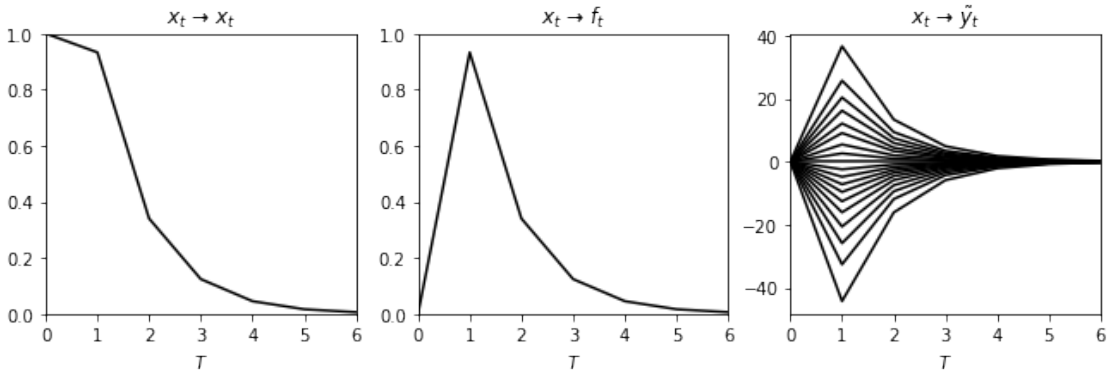


Figure 10: Impulse response functions for unit shock in differenced unemployment rate x_t on (from left to right) x_t, common factor f_t and differenced transitions into education $\tilde{y}_t$

6 Conclusion

In this paper we have introduced a novel dynamic factor model that is capable of forecasting the number of students across the many different types of education. The model can also be used to analyze the relationship between education participation and relevant macroeconomic variables such as the unemployment rate. We further have proposed an econometric treatment for this flexible modeling framework. The empirical analysis is carried out for a large dataset for the educational system in the Netherlands. In this study we have found that, overall, changes in the unemployment rate account for approximately 7.7% of the changes in the flows across the educational system. Given that the panel data dimension is huge, we have allowed for clustering in the factor loadings that are associated with the dynamic macroeconomic factor. As a result we have been able to measure the extent in which the different types of education exhibit similarities in their relationship with macroeconomic cycles. The Monte Carlo study has shown that our developed methodology is able to correctly identify clusters, while the empirical analysis has highlighted its practical feasibility and its good forecasting performance. In future research, we plan to generalize the methodology further and verify its theoretical properties.

References

Alonso, A. M., Galeano, P., & Peña, D. (2020). A robust procedure to build dynamic factor models with cluster structure. *Journal of Econometrics*, *216*(1), 35-52.

Bai, J., & Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, *70*(1), 191–221.

Barnichon, R., & Mesters, G. (2018). On the demographic adjustment of unemployment. *Review of Economic Statistics*, *100*(2), 219–231.

Blasques, F., Koopman, S. J., & Lucas, A. (2016). *Maximum likelihood estimation for generalized autoregressive score models* (Tech. Rep.). VU Amsterdam.

Bräuning, F., & Koopman, S. J. (2014). Forecasting macroeconomic variables using collapsed dynamic factor analysis. *International Journal of Forecasting*, *30*(3), 572-584.

Bureau for Economic Policy Analysis. (2019, November). *MLT-raming november 2019, cijfers.* Retrieved 2020-05-21, from https://www.cpb.nl/middellangetermijnverkenning-2022-2025#docid-160027

Chen, X., & Liao, Z. (2014). Sieve M inference on irregular parameters. *Journal of Econometrics*, *182*, 70–86.

Clark, D. (2011). Do recessions keep students in school? The impact of youth unemployment on enrolment in post-compulsory education in England. *Economica*, *78*(311), 523–545. (Publisher: Wiley Online Library)

Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, *28*(5), 777-795.

Doz, C., Giannone, D., & Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on kalman filtering. *Journal of Econometrics*, *164*(1), 188-205.

Groenez, S., Desmedt, E., & Nicaise, I. (2007). Participation in lifelong learning in the eu-15: the role of macro-level determinants. In *Paper for the ecer conference.*

Hallin, M., & Liška, R. (2011). Dynamic factors in the presence of blocks. *Journal of Econometrics*,

163, 29–41.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series* (Vol. 52). Cambridge University Press.

Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data mining and knowledge discovery*, *2*(3), 283–304.

Jungbacker, B., & Koopman, S. J. (2015). Likelihood-based Dynamic Factor Analysis for Measurement and Forecasting. *The Econometrics Journal*, *18*(2), C1–C21.

Lamb, S., Walstab, A., Teese, R., Vickers, M., & Rumberger, R. (2004). Staying on at school: Improving student retention in australia. *Brisbane: Queensland Department of Education and the Arts*.

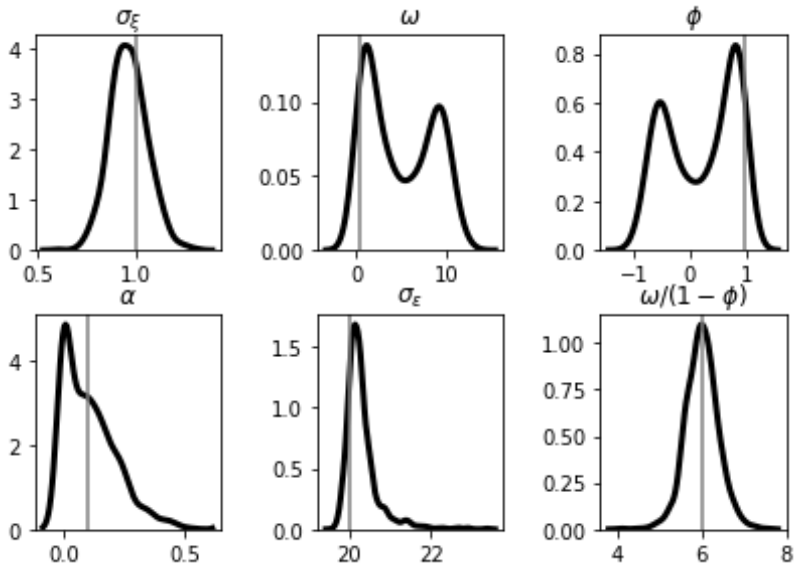
Ministry of Education, Culture and Science. (2020, June). *Referentieramingen 2020.*

Spijkerman, M. (2006, October). *De invloed van conjunctuureffecten op onderwijsdeelname.* SEOR, Erasmus Universiteit Rotterdam.

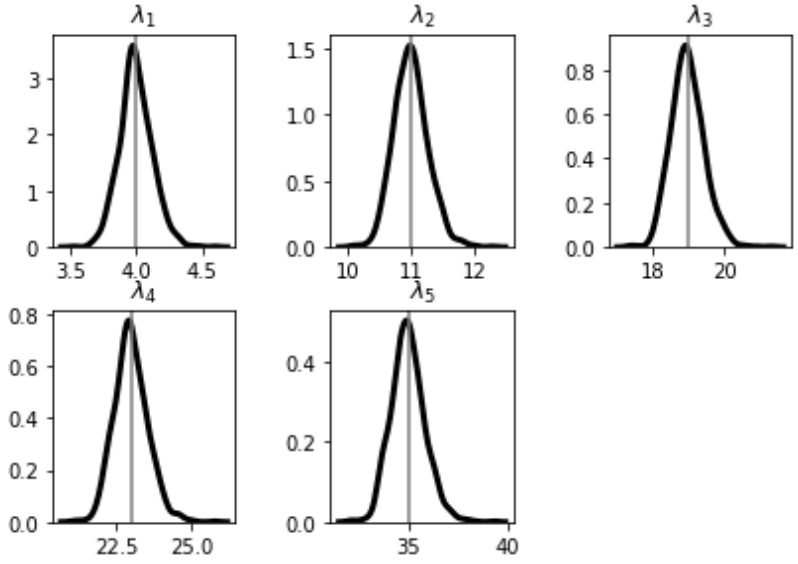
Stock, J. H., & Watson, M. W. (2002). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179.

Stock, J. H., & Watson, M. W. (2008). The evolution of national and regional factors in u.s. housing construction. In T. Bollerslev, J. Russell, & M. Watson (Eds.), *Volatility and time series econometrics: Essays in honor of Robert F. Engle.* Oxford University Press.

A Appendix: Additional Monte Carlo and Empirical Results

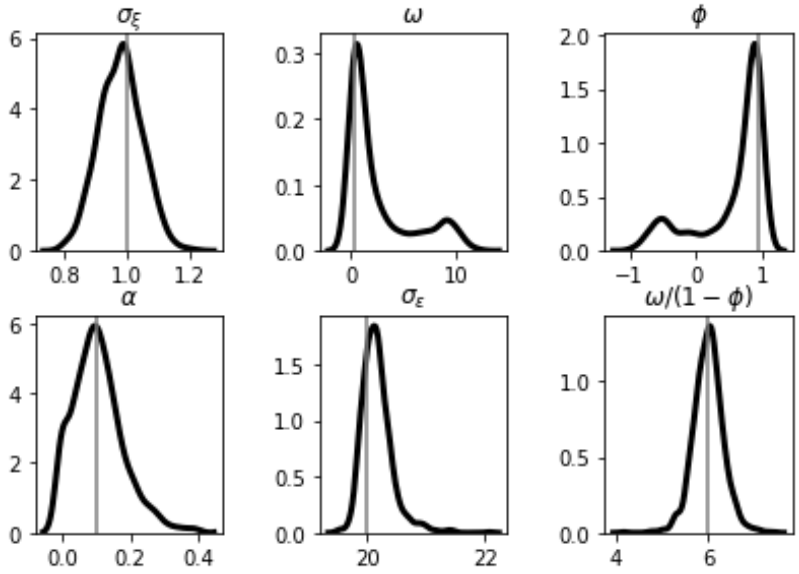


(a) Densities of estimated static parameters.

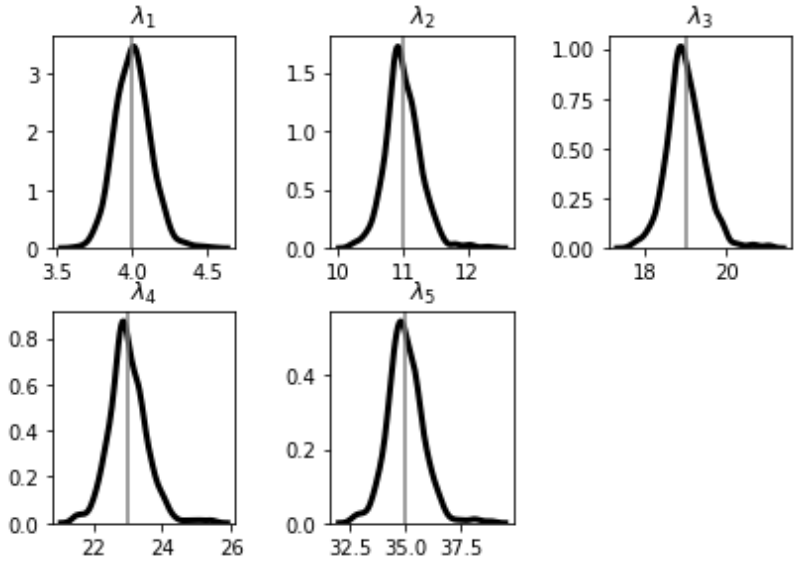


(b) Densities of estimated cluster centroids.

Figure A.1: Parameter estimation results of $M = 1,000$ simulations for $N = 500, T_y = 20, T_x = 50$. Vertical lines represent true values.



(a) Densities of estimated static parameters.



(b) Densities of estimated cluster centroids.

Figure A.2: Parameter estimation results of $M = 1,000$ simulations for $N = 500, T_y = 20, T_x = 100$. Vertical lines represent true values.

$\mathbf{T_y = 20, T_x = 50}$

	MSE			MAE			Unrestricted		Clustered		
	Unr.	**Cl.**	**Diff.**	**Unr.**	**Cl.**	**Diff.**	**LL**	**AIC**	**LL**	**AIC**	**LR**
C1	0.570	0.016	0.555	0.602	0.097	0.594	-8,694	17,590	-8,745	17,494	102
C2	0.629	0.076	0.554	0.631	0.213	0.593	-8,702	17,606	-8,751	17,506	98
C3	0.769	0.284	0.548	0.697	0.375	0.592	-8,717	17,636	-8,765	17,534	96
C4	0.865	0.372	0.544	0.738	0.448	0.590	-8,727	17,656	-8,775	17,554	96
C5	1.277	0.722	0.555	0.891	0.657	0.593	-8,767	17,736	-8,814	17,632	94
Full	0.822	0.294	0.551	0.712	0.358	0.592	-44,048	89,098	-44,293	88,598	490

$\mathbf{T_y = 20, T_x = 100}$

	MSE			MAE			Unrestricted		Clustered		
	Unr.	**Cl.**	**Diff.**	**Unr.**	**Cl.**	**Diff.**	**LL**	**AIC**	**LL**	**AIC**	**LR**
C1	0.571	0.014	0.557	0.603	0.093	0.595	-8,694	17,590	-8,745	17,494	102
C2	0.630	0.075	0.555	0.631	0.207	0.593	-8,697	17,596	-8,747	17,498	100
C3	0.765	0.276	0.547	0.691	0.361	0.590	-8,706	17,614	-8,755	17,514	98
C4	0.851	0.358	0.544	0.727	0.426	0.589	-8,709	17,620	-8,758	17,520	98
C5	1.240	0.686	0.554	0.869	0.623	0.592	-8,735	17,672	-8,784	17,572	98
Full	0.812	0.282	0.552	0.704	0.342	0.592	-43,982	88,966	-44,230	88,472	496

Table A.1: Model fit for unrestricted and clustered model. All results are based on $M = 1,000$ simulations with $N = 500$ and $T_y = 20$, with $T_x = 50$ in the top panel and $T_x = 100$ in the bottom panel. Each row in a panel represents a cluster and the last row is the full sample. In the columns are loss functions given of the unrestricted loadings compared to the true ones ("Unr."), the clustered centroids compared to the true ones ("Cl.") and the unrestricted loadings minus the clustered centroids ("Diff."). The first three columns have the MSE as loss function and the last three columns the MAE. For both models, the log likelihood and AIC are given and they are compared via the LR-statistic in the last column. For the latter, the critical values are 123 for each cluster (100-1 degrees of freedom) and 548 overall (500-5 degrees of freedom), at the 5% significance level.

		-47.5	-35.0	-27.8	-22.1	-17.3	-13.5	-10.3	-7.7	-5.2
hbo	fulltime	1	2	7	8	17	18	13	33	30
	parttime	4	2	3	8	14	4	9	9	8
mbo	school-based	0	2	1	4	1	6	5	4	11
	on-the-job	6	13	17	18	16	14	11	12	8
wo	fulltime	1	1	6	9	18	31	25	29	30
	parttime	0	0	0	3	2	4	2	2	1

		-2.6	0.0	2.8	5.8	9.7	13.0	17.3	21.8	27.5	39.2
hbo	fulltime	32	33	40	36	20	10	9	3	2	0
	parttime	10	2	5	4	3	3	0	0	2	2
mbo	school-based	13	26	23	27	31	28	18	15	19	8
	on-the-job	11	9	5	15	3	4	6	5	4	0
wo	fulltime	37	47	36	18	10	3	3	1	1	0
	parttime	1	4	1	3	0	0	0	0	0	1

Table A.2: Clusters (centroid position in column names) with composition of education types. Translations: *mbo*: vocational education, *hbo*: hbo, *wo*: university