Forecasting internal migration in the short to medium run

Trond Husby*and Hans Visser*

*Netherlands Environmental Assessment Agency (PBL). P.O. box 30314, 2500 GH, the Hague, the Netherlands. Corresponding author trond.husby@pbl.nl

Abstract

BACKGROUND
Long-term projections of annual internal migration are key inputs to subnational population projections. These long-term projections are commonly extrapolation of a long-term trend in the data. Since time series of internal migration follow economic cycles it is informative to analyse its short- to medium-term dynamics using data of monthly frequency.
OBJECTIVE
This paper develops a model for short-to medium term univariate forecast of monthly frequency of mobility in the Netherlands. Of particular interest is a forecast around the changepoint of a cycle.
METHOD
The model is a dynamic linear model, which belong to the state space family of models, incorporating trend, seasonal and autoregressive components. The developed model allows to trade-off between short- and long-run trends in the data.
RESULTS
Forecast accuracy is evaluated using time series cross-validation (evaluation over a rolling horizon). Forecast errors are compared to those of several other popular univariate forecasting models. The model developed in the paper is competitive or even better than the models included as comparison.
CONCLUSION
The analysis reveals that the model accurately forecasts the frequency of mobility at least a year ahead.
CONTRIBUTION
The paper shows how short- to medium-term forecasts of monthly data can be used to inform long-term projections as used in subnational population projections.
This version: February 19, 2020

1 Introduction

The future development of migration is a key input to subnational population projections. It is also a major source of uncertainty: due to strong fluctuations of migration flows, a large part of the uncertainty in the projected population can be attributed to migration (Beer 1993), both external and internal. Long-run forecasts of internal migration, such as those used in subnational population projections, are usually based on extrapolations

of a long-term trend, and are almost always based on annual data. However, trend extrapolation is problematic if short- to-medium run fluctuations stretch into the long run or if the most recent data points in the time series represent the top or the bottom of a cycle (Canova 1998; Hamilton 2018). Of course, one of the problems with annual time series is that they are released once a year, and the top of a cycle may be somewhere in the middle of that year. In this paper we show how an analysis of the short- to medium-term dynamics of internal migration on monthly data can provide early warning signals about possible changes in the trend in the annual data.

As is the case in most developed countries, the main driver of regional-level population change in the Netherlands is migration (external and internal) (te Riele et al. 2019). Due to the close relation between migration decisions and labour- and housing market conditions, the volume of internal migration often moves along with the macroeconomic cycle (Husby, Weterings, and de Groot 2019). However, recent research argues that the relationship between internal migration and labour market conditions has changed over time (Kaplan and Schulhofer-Wohl 2017). Moreover, other factors than macroeconomic drivers may, at least in the long run, be just as important in explaining migration. For example, family considerations are important determinants of migration decisions on an individual level, meaning migration flows may well be affected by changes to family composition and aging (Mulder 2018). As such, there are many candidate determinants of internal migration and their relationship with internal migration flows may change over time. Therefore it is questionable whether the use of explanatory variables will improve the forecast accuracy (Makridakis, Hyndman, and Petropoulos 2019).

Even if one succeeds in finding such explanatory variables, it is far from certain that their inclusion improves the accuracy of the population projections (Smith 1997). Many official population projections are carried out with cohort-component models, where the growth paths of the components are given as inputs to the model (see e.g., de Jong et al. 2005). A forecast of migration using explanatory variables requires knowledge of both the future trajectory of the explanatory variables and the future development of their relationship with migration. Consequently, the growth path used as inputs in cohort-component models is in morst practical cases based on extrapolation of long-term historic trends (Smith, Tayman, and Swanson 2013).

When working with the latest Dutch Regional Population Projections, published in September 2019 (te Riele et al. 2019), we were asking ourselves whether the last data point in the time series of annual frequency of mobility, namely end of 2016, represented the top of a cycle. We knew, for example, that a recent change to the Dutch student finance system may have caused

a decrease in mobility for an otherwise very mobile group.[1] Obviously, if the last data points used to determine the long-term trend represent the top of a cycle, a trend extrapolation could overestimate the true long-term trend. This was the direct motivation behind this paper: could a short- to medium-term forecast of monthly data provide more information about the future development of internal migration on annual data?

Univariate forecasting of time series involves decomposing the series into elements such as trend, cycle and seasonality, where these basic time series patterns are used to form a forecast (Zietz and Traian 2014). One model type frequenctly used for such analyses is Dynamic Linear Models (DLM) which belong to the familiy of state space models (Petris, Petrone, and Campagnoli 2009; Durbin and Koopman 2012). DLMs are linear models with a Gaussian error structure, where the relevant inferences are carried out using the Kalman algorithm. One advantage of this model type is that the stochastic nature of the Kalman algorithm allows for derivation of forecast distributions. Another advantage is that these models are easily extended: time series components can be added to or removed from the model if deemed necessary.

In this paper we develop a dynamic linear model for a short- to medium-run forecast of the frequency of migration, where short-to medium run is understood as a time horizon up to two years. We compare the forecasting performance on our model with that of four other popular models for univariate time-series forecasting. These models used for comparison are all fit using automatic routines available in *R*. As such, as a byproduct of the evaluation exercise, we evaluate the merits of our manual model selection compared to the automatic model selection in easy-to-use software packages. The code used for the analysis in the paper is entirely self contained and is publicly available at `https://github.com/trondhusby/mobility_forecast`.

The structure of the paper is as follows: section 2 shows the time series data and presents the model; section 3 shows the estimated parameters and results from basic inference with the; section 4 presents results from an out-of-sample evaluation and illustrates how forecasts of monthly can inform future development on an annual scale; section 5 concludes.

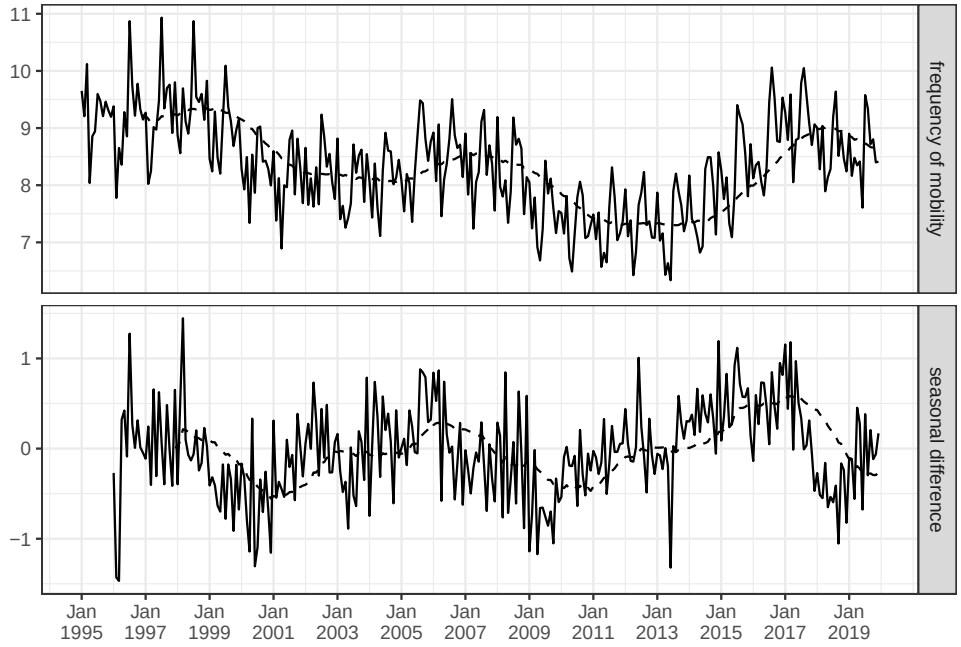
[1] https://www.cbs.nl/nl-nl/nieuws/2018/04/studenten-gaan-minder-op-kamers

2 Data and modelling

2.1 Monthly time series of the frequency of mobility

The time-series in this paper is the national-level frequency of mobility, defined as the sum of intra- and inter-municipal moves per 1000 inhabitants. Open data available from the Statistics Netherlands allow us to calculate monthly time series of the from January 1995 to December 2019, meaning we have in total 300 observations at our disposal. More formally, we define the frequency of mobility in month t as the sum of inter- en intra-municipality moves between the first and the last day of month t, divided by the population on the last day of t. Figure 1 shows the time series from January 1995 to December 2019 as well as the seasonal differences. The dotted line in the figure shows two-year moving averages of both the frequency of mobility and the seasonal differences.

Figure 1: Frequency of mobility (upper panel) and seasonal differences (lower panel) and two-year moving averages (dotted line)



The upper panel shows that there are between 11 and 7 moves per 1000 inhabitants per month. The plot suggests that the frequency of mobility is cyclical with uneven period length. We see clearly from the plot that the time series is non-stationary and a Dickey Fuller test confirms this. The financial

crisis, which hit the Dutch housing market hard, can be seen in the dip from 2008 onwards, with the recovery from the crisis set in during the year of 2014. The dotted line, which in this panel can be interpreted as the trend-cycle, illustrates clearly the connection between mobility and the macroeconomy: the dip from 2008 onwards and the recovery from 2014 follows the financial crisis which hit both labour and housing markets in the same periods.

From the bottom panel, which shows the seasonal differences, we see that the year on year changes are negative from January 2009 with a recovery in 2012 and a new dip in 2013 before a real recovery from 2014. Interestingly the lower-panel figure reveals that the year on year changes become negative from medio 2017. This is also reported by Statistics Netherlands:[2] in 2018 there were 5 % fewer moves than in 2017, with the decline concentrated primarily among people younger than 50 years old. This reflects thus an end of an increasing trend from 2014 onwards. As such, our suspicions of a peak in mobility frequency around the end of 2016 seem to be justified.

Summing up, the decomposition of the time series suggests that there is no linear trend, rather a cycle with peaks roughly every 10 years. There are furthermore strong seasonal effects with variations between the years, meaning the seasonality contains noise. Furthermore, the seasonality and the trend-cycle lead to autocorrelation. The next section shows how we adress these issues using the state space model.

2.2 Model description

The state space model is a representation of a dynamic system where observed values are a linear function of an unobserved process (the state) and noise. In other words, if y_t is an observed time series and v_t is measurement error, then the true underlying process of y_t is given by the vector θ_t. The state space formulation is necessary to carry out inference using the Kalman filter algorithm (Kalman 1960), which is a computational recursive means to estimate the state θ_t in a way that minimizes the mean of the squared error. The filter supports estimations of past, present, and even future states. As mentioned above, DLMs form a special case of state space models where errors are normally distributed and indepentently distributed. There are many ways of describing DLMs; the model is implemented in *R* using the *dlm* package, therefore the description in this section follows the notation used in Petris, Petrone, and Campagnoli (2009).

Let y_t denote the logarithm of the frequency of mobility in month t, which are the observed values of underlying (unobserved) vector of states θ_t. y_t are

[2] `https://www.cbs.nl/nl-nl/nieuws/2018/26/minder-mensen-verhuisd-in-eerste-kwartaal-2018`

conditionally independent given the state θ_t, and the state is a latent Markov process, meaning that the probability of moving to the next state depends only on the previous state. We can write this DLM as:

$$y_t = F\theta_t + v_t \qquad v_t \sim N(0, V_t) \tag{1}$$
$$\theta_t = G\theta_{t-1} + w_t \qquad w_t \sim N(0, W_t) \tag{2}$$
$$\theta_0 \sim N(m_0, C_0) \tag{3}$$

The first equation is called the observation equation and the second the state equation. v_t and w_t are uncorrelated Gaussian errors, where the observation variances are gathered in the $m \times m$ matrix V_t and system variances in the $p \times p$ matrix W_t. F and G are known system matrices of dimension $p \times m$ and $p \times p$ respectively. The initial θ_0 are normally distributed with means m_0 and variances C_0.

Elements of the state vector θ_t should be chosen such that they reflect the characteristics of the time series. One thing we noted in the previous section was that periods of growth were followed by periods of decline, meaning there were piecewise or local linear trends. One model that can capture such patterns is the linear growth or local linear trend model, where the y_t are noisy observations of a level μ_t which varies over time with slope β_t. The dynamics of the slope itself is modelled as a random walk. We can write the local linear trend model as follows:

$$y_t = \mu_t + v_t \qquad v_t \sim N(0, \sigma_v^2) \tag{4}$$
$$\mu_t = \mu_{t-1} + \beta_{t-1} + w_{t,1} \qquad w_{t,1} \sim N(0, \sigma_\mu^2) \tag{5}$$
$$\beta_t = \beta_{t-1} + w_{t,2} \qquad w_{t,2} \sim N(0, \sigma_\beta^2) \tag{6}$$

Following the notation in Petris, Petrone, and Campagnoli (2009), we can write the local linear trend model as:

$$\theta_t = \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} \quad G_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad F_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \quad W = \begin{bmatrix} \sigma_\mu^2 & 0 \\ 0 & \sigma_\beta^2 \end{bmatrix} \quad V = \begin{bmatrix} \sigma_v^2 \end{bmatrix}$$

Seasonality is dealt with by including a trigonometric seasonal components to the model with a fixed periodicity of 12 months. One advantage with the trigonometric specification, relative to simpler seasonal dummies, is that we allow autocorrelation to last through more lags, resulting in a smoother seasonal pattern. This way we filter out some of the noise seen in

the figures in the previous section. By including non-zero variances the seasonal pattern is allowed to change trough time. Seasonality is incorporated in the DLM model by extending the state equation with a trigonometric function that describes a harmonic motion. For the jth harmonic we can write the evolution of the seasonal effects as:

$$S_{j,t+1} = S_{j,t}\cos\omega_j + S^*_{j,t}\sin\omega_j \tag{7}$$
$$S^*_{j,t+1} = -S_{j,t}\sin\omega_j + S^*_{j,t}\cos\omega_j \tag{8}$$

with the fourier frequencies $\omega_j = \frac{2\pi tj}{s}$. We include two harmonics, meaning that the trigonometric specification can be written on matrix-form as:

$$\theta_t = \begin{bmatrix} S_{1,t} \\ S^*_{1,t} \\ S_{2,t} \\ S^*_{2,t} \end{bmatrix} \qquad G = \begin{bmatrix} \cos\omega_1 & \sin\omega_1 & 0 & 0 \\ -\sin\omega_1 & \cos\omega_1 & 0 & 0 \\ 0 & 0 & \cos\omega_2 & \sin\omega_2 \\ 0 & 0 & -\sin\omega_2 & \cos\omega_2 \end{bmatrix}$$

$$F = \begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix} \qquad W = \begin{bmatrix} \sigma^2_{S_1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{S_2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The two components discussed so far (trend and seasonal) remove much of the autocorrelation, however a Ljung-Box test confirmed that the residuals of such a model were still not white noise. One way of dealing with residual autocorrelation is to include autoregressive elements into the state space model. The specification chosen here is an AR(4) with two higher order lags, namely lags 7 and 12. This model can be written on matrix notation using the following trick:

$$\begin{bmatrix} \mu_{1,t} \\ \mu_{2,t} \\ \mu_{7,t} \\ \mu_{12,t} \end{bmatrix} = \begin{bmatrix} \phi_1 & 1 & 0 & 0 \\ \phi_2 & 0 & 1 & 0 \\ \phi_7 & 0 & 0 & 1 \\ \phi_{12} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu_{1,t-1} \\ \mu_{2,t-1} \\ \mu_{7,t-1} \\ \mu_{12,t-1} \end{bmatrix} \begin{bmatrix} u_t \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

which gives

$$\mu_{1,t} = \phi_1\mu_{1,t-1} + \phi_2\mu_{1,t-2} + \phi_7\mu_{1,t-7} + \phi_{12}\mu_{1,t-12} + u_t \quad u \sim N(0, \sigma^2_u) \tag{9}$$

Note that the standard routine for creating autoregressive elements in the *dlm* package does not allow for 'gaps' in the included lags. In order to estimate the model, we use the system matrices of an AR(12) where the parameter of lags not included in the model are set to zero:

$$G = \begin{bmatrix} \phi_1 & 1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \phi_7 & 0 & 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 1 \\ \phi_{12} & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \quad F = \begin{bmatrix} 1 & 0 \cdots \end{bmatrix} \quad W = diag(\sigma_u^2, 0, ...)$$

3 Results

Before we apply the Kalman filter for estimation and forecasting we first need to determine the unknown model parameters. We chose to estimate all of them using maximum likelihood (see Petris, Petrone, and Campagnoli 2009, ch. 4). With the maximum likelihood estimate of the unknown parameters we can use the Kalman filter for inference. One important inferential tasks is to estimate θ_t with data up to t. This is referred to as filtering. Another closely related task is smoothing, which entails estimating θ_t with all data. This section presents the estimated parameters and the resulting filtering and smoothing estimates of the frequency of mobility.

3.1 Estimated parameters

As we saw in the previous section, the matrices W and G contain 10 potentially non-zero parameters: σ_v^2, σ_μ^2, σ_β^2, $\sigma_{S_1}^2$, $\sigma_{S_2}^2$, σ_u^2, ϕ_1, ϕ_2, ϕ_7, and ϕ_{12}. Note that the parameters for the autoregressive component ϕ_1, ϕ_2, ϕ_7, and ϕ_{12} are estimated subject to stationarity restrictions (Monahan 1984). Initial analyses revealed that the value of $\sigma_{S_1}^2$ discernible impacts on model results, hence this parameter is simply set to zero. The first two parameters in W represent the variances for the trend-cycle component: one for the level and one for the slope. Initial runs with the model revealed that two special cases fit the data better than the general formulation with both variances larger than zero. The first special case is obtained by setting $\sigma_\mu^2 = 0$, sometimes referred to as the integrated random walk or the smooth trend model (Young

et al. 1991). In this specification, all noise is shifted from the level component to the slope, meaning that the trend-cycle component of the state vector becomes rather smooth (hence the name). In the rest of the paper we will refer to this model as DLM1. A second special case is a model in which $\sigma_\beta^2 = 0$, where all stochasticity is removed from the dynamics of the slope.In this formulation the estimated β_t reflect the long run slope of the time series rather than the local slope, essentially reducing to what is sometimes called the local trend model (Durbin and Koopman 2012). Due to the inclusion of the autoregressive elements, the forecasted values can diverge from the long-term growth in the short-run. More specifically, a one-step-ahead forecast $\hat{y}_{t+1}$ reflects up to 12 lagged values of the time series y_t and not only the long-term trend. We will refer to this model as DLM2.

The estimated parameters are shown in Table 1. We see that the observation variance σ_v^2 is very close to zero, indicating a high level of precision of the observations. The estimated parameters of the autoregressive parts are very similar between the two models, except for the parameter of the second lag which takes positive values for DLM1 and negative values for DLM2. $\sigma_{S_2}^2$ is larger than zero for both models, capturing changes in the seasonal component.

Table 1: Estimated variances and AR parameters

Parameter	DLM1	DLM2
σ_v^2	1.523e-08	1.523e-08
σ_μ^2	0	0.0001457
σ_β^2	2.511e-06	0
$\sigma_{S_2}^2$	7.721e-06	7.727e-06
σ_u^2	0.001284	0.001022
ϕ_1	-0.643	-0.6951
ϕ_2	0.03487	-0.05413
ϕ_7	-0.4429	-0.4896
ϕ_{12}	0.4511	0.4358

3.2 Filtering and smoothing estimates of the frequency of mobility

We obtain the filtered distribution of θ_t conditional on the observed series up to t, $\theta_t|y_1, y_2, ..., y_t$. Forecasts of the next observation y_{t+1} based on observations up to t, are produced by first computing the state vector θ_{t+1} and then predicting $\hat{y}_{t+1}$. Similarly, an n-step ahead forecast y_{t+n} is based on calculat-

ing the n-step ahead state vector θ_{t+n}. We can write the forecast function as $f_t = E(y_{t+n}|y_1, y_2, ...y_t)$. In addition we estimate a smoothing distribution of the components, representing the past values of the states given all observed values $\theta_t|y_1, y_2, ..., y_s$ where $s \geq t$.

The one-step ahead forecast from the Kalman filter, or filtering estimates, from both of the models are displayed in Figure 2 together with Mean Absolute Percentage Error (MAPE). Due to the recursive nature of the Kalman algorithm, the filtered values in the first periods veer quite far off from the observed values (not shown in the figure). We therefore discard the three first years when calculating the MAPE. The MAPE suggests that DLM2 fits the data somewhat better than DLM1, though the difference is not dramatic. A residual check showed that the normality assumption of DLMs is justified and that residuals are essentially white noise (see Appendix B).

Figure 2: One-step-ahead forecast (solid line), observed data (dotted line) and MAPE (calculated from January 1998)

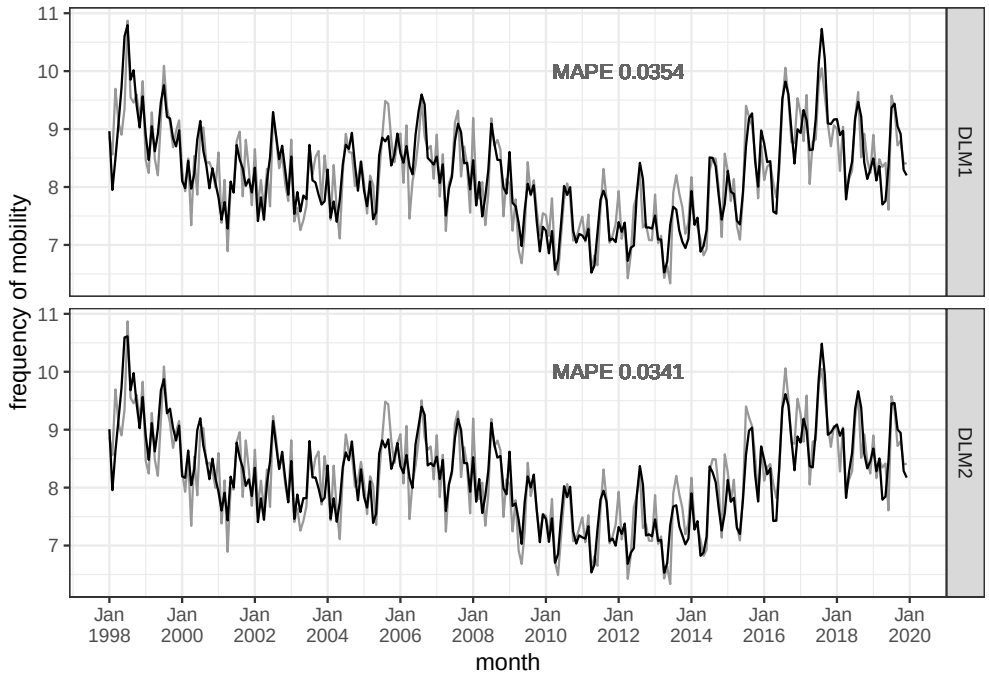
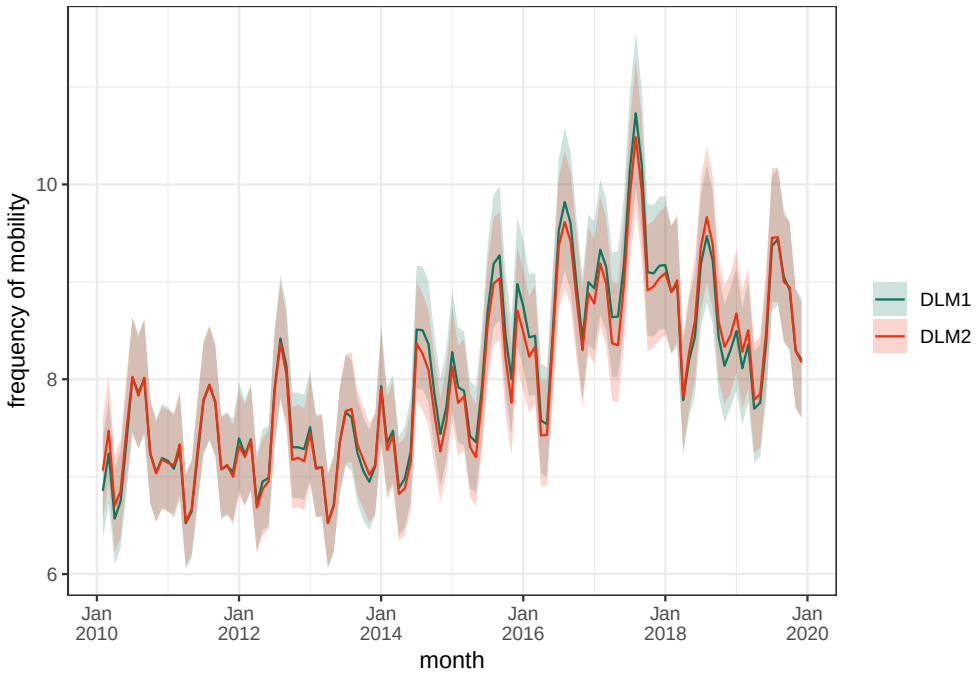


Figure 3 shows the one-step ahead forecasts along with the 95 percent prediction intervals of DLM1 and DLM2 for the months between January 2010 and December 2018. The interval is calculated using the standard deviation of the filtered values (Petris, Petrone, and Campagnoli 2009, ch. 3). Although the mean forecasts of the two models are in general quite similar,

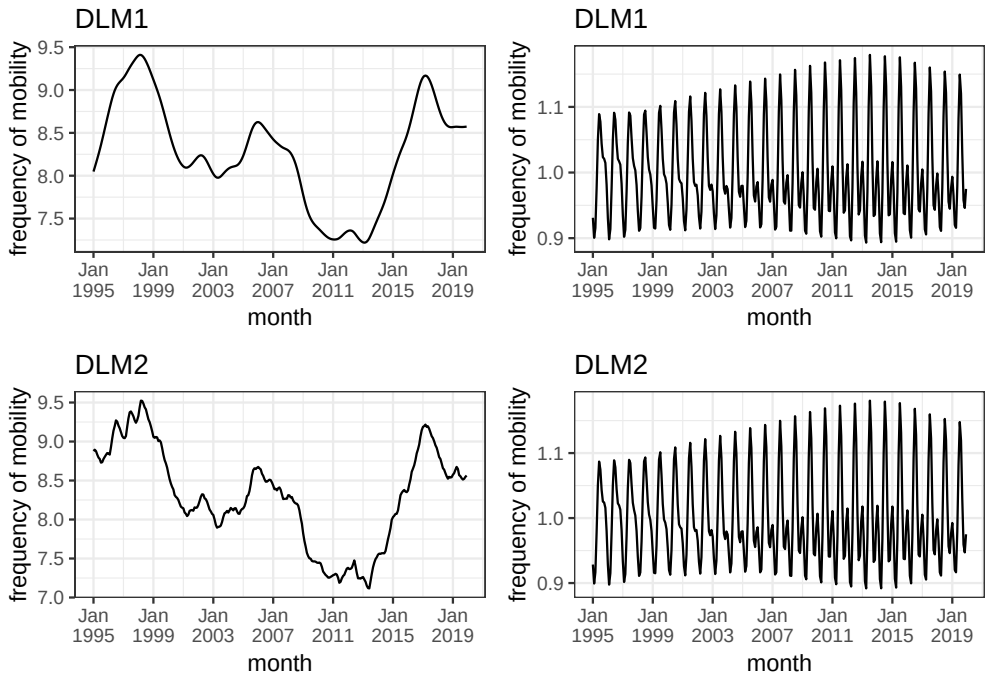
we see that DLM1 systematically predicts a higher frequency of mobility than DLM2 from 2014 to about 2017.

Figure 3: Filtering estimate and 95 percent prediction intervals



The smoothing estimates for the trend-cycle component and for the seasonal effects are shown in Figure 4. From the left panels we can pinpoint the recent peak in the trend-cycle to March 2017. We see a visible difference between DLM1 and DLM2, where the former model produces a much smoother trend-cycle. The right panels show how the seasonal effects vary over time; at the start of the time series the within-year cycle exhibites one pronounced maximum (July) and one minimum (March). However, from around the middle of the series, there is gradually another local maximum within each year (January). This means that the seasonal patterns in the model captures well the development of the seasonal factors from the raw data.

Figure 4: Smoothing estimates of the trend-cycle (left) and seasonality (right)



4 Discussion

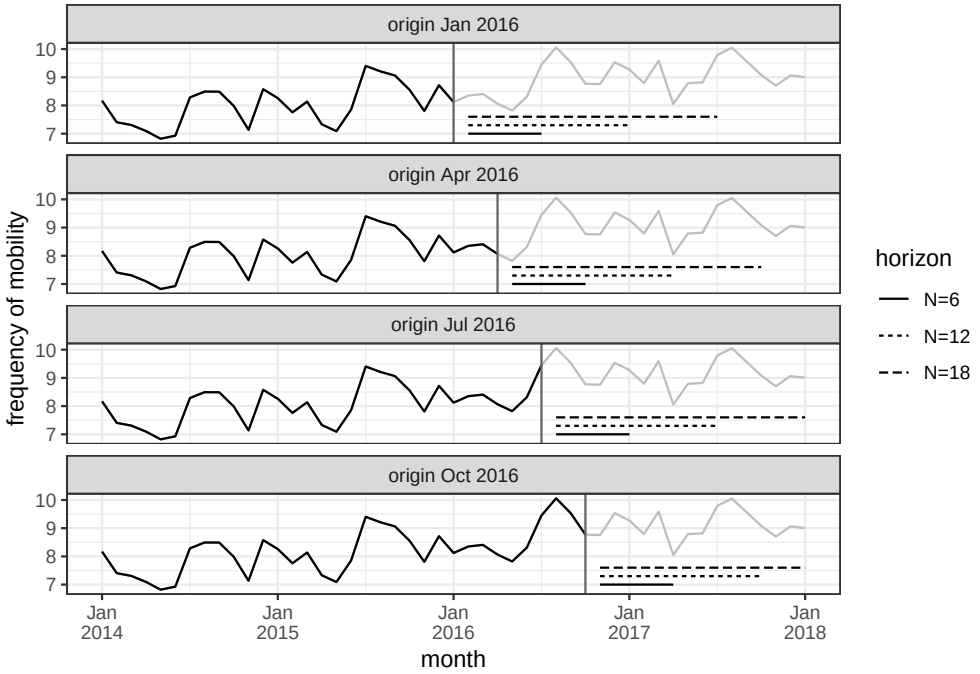
In this section we evaluate the forecasting accuracy of the models described in the previous sections. Our primary interests here are how well the models predict data points that have not been used for estimation (out-of-sample), and how performance varies with the length of the forecast horizon, whereby accuracy measures are calculated for several forecast lengths (in effect N-step forecast).

4.1 Cross validation: evaluation over a rolling origin

For cross-sectional data model evaluations are commonly carried out by estimating the model on a training data set and evaluating it on an independent test data set. In order to control for effects arising from the composition of the training data, it is common to repeat this procedure with different training and test data sets. This procedue is called k-fold cross validation. For time series it is common to choose a test set that does not contain observations occuring prior to the observations in the training data (Bergmeir and Benıtez 2012). Recent literature suggests that standard cross-validation can

be applied to certain time series models (Bergmeir, Hyndman, and Koo 2018), however, due the Markov assumption embedded in the Kalman filtering algorithm, we suspect that the technique would not work for the the DLMs developed in this paper. Therefore we follow the conventional approach for evaluating forecast errors; namely evaluation over a rolling forecasting origin (Tashman 2000) where the training data is successively extended in k iterations. The last point of the training data in each iteration is referred to as the origin T, since the origin on which the forecast is based rolls forward in time. In this approach, the training data consists of observations between $t = 1, ..., T$ and forecasts are generated for time periods $T + 1$, $T + 2$, ..., $T + N$. This procedure is illustrated in Figure 5. The movement along the x-axis shows how the origins (vertical dotted line) "roll" forward in time, where the model is estimated on the training data (black solid line). Forecast perfomance is then evaluated by averaging the forecast errors over different horizons (dotted horizontal lines).

Figure 5: Illustration of rolling forecasting origin: training data (solid black line), origin (vertical dark grey line), forecast horizion (dotted horizontal lines) and test data (grey line)



We evaluate forecasting performance on three time horizons: $N = 6$, $N = 12$ and $N = 18$. This means that, for each origin, we calculate three indicators of accuracy; one for a half year forecast, one for a year and one for

one and a half year. We are interested in assessing the performance around the changepoint of the trend-cycle. From the previous section we identified this as March 2017 (see Figure 4). In order to evaluate the performance of the models on both sides of the changepoint we include origins within 12 months before and after March 2017. Including the changepoint itself we have in total 25 forecast origins: T = March 2016, February 2016, ..., June 2017.

It is possible that estimated parameter values differ between the different origins. The resulting mean and standard deviations, shown in Table 3 in the Appendix, reveal that differences in training data has a very limited effect on the estimated parameters. This also gives us confidence that there is enough data for estimation even on the first origin.

Performance of the DLMs is assessed by comparing forecasting accuracy with that of six other popular models. The first of these is a naïve seasonal model (Naïve), where the forecast of a specific month is simply the value of the same month of the previous year. Despite its simplicity this model often performs very well, especially for economic and financial time series.

The next model is the Holt-Winters method with multiplicative seasonality (Holt 2004; Winters 1960). This model is an extension of the simple exponential smoothing model, allowing for forecasts with seasonality. Using the function `HoltWinters`() in base *R*, the smoothing parameters of the model are selected automatically, for each origin, with the default initial values. This model type performs particularly well on data with a clear trend, and we therefore expect it to forecast accurately before the changepoint and increasingly worse around and after it.

Furthermore, exponential smoothing can also be formulated on state space form, which in this sense becomes a general formulation of exponential smoothing methods. We will refer to this model type as ETS (Error, Trend, Seasonal). One advantage of formulating exponential smoothing models on the (stochastic) state space form is the possibility of generating prediction intervals - either analytically or by simulation. This furthermore allows information criteria to be used for model selection, meaning that all combination of trend-, seasonal and error components can be explored (Hyndman et al. 2002). The automatic model selection is carried out using the `ets`() function with default values for all arguments, meaning that selection is based on the corrected Akaike's Information Criterion (AICc).

The Structural Time Series approach, as described in Harvey (1990), is an alternative way of writing and estimating state space models. The function `StructTS`() available in base *R* can be used to estimate some simple variants of this model type. We use this function to estimate a local linear growth model with simple dummies to control for seasonality.

Another widely used model, the so-called TBATS model, extends the ETS model by allowing for more complicated representations of sesonality (De Livera, Hyndman, and Snyder 2011). It allows for the Fourier-specification of seasonality that we used in both of the DLMs. Also the TBATS model can be automatically specified using a similar routine as that of the ETS model.

Finally we include a seasonal ARIMA model into the comparison. Identification of the model is carried out using the `auto.arima()` function on the whole sample (Hyndman and Khandakar 2008), and results in an ARIMA(2,1,5)(2,1,1). This is thus a model with double differencing, with yearly and monthly MA(2), and yearly AR(5) and monthly AR(2). For each origin we reestimate the model parameters on the respective training data. We also tried the automatic selection for each origin with only marginal differences in results.

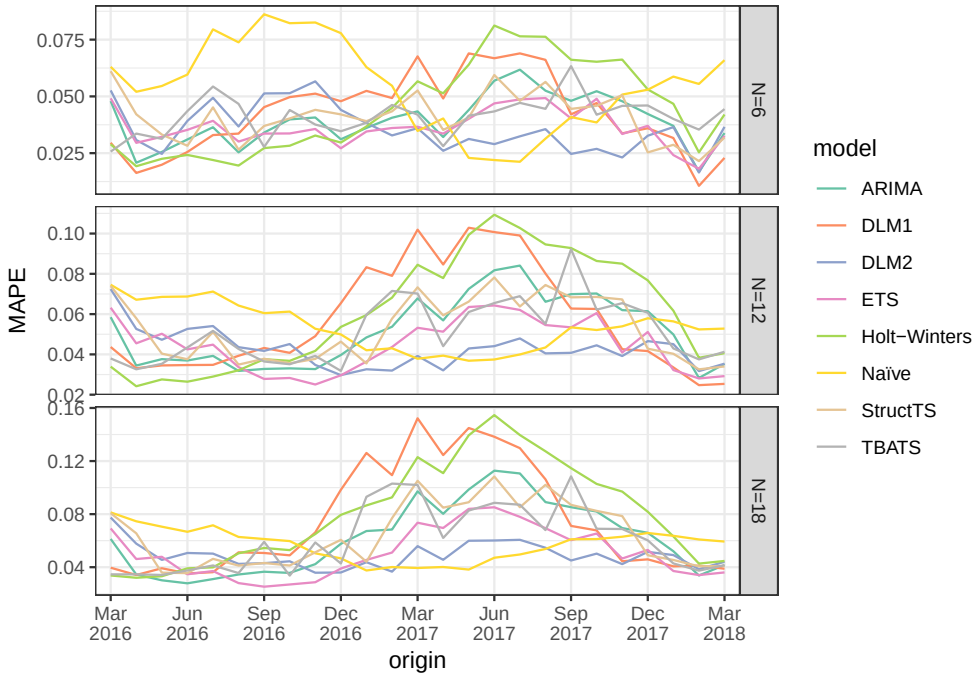
Table 2 shows the overall forecast accuracy across origins. It shows both MAPE and root mean square error (RMSE), averaged over origins, for each forecast horizon (6 months, 12 months, 18 months). The standard deviation of RMSE and MAPE, which is a measure of how much accuracy varies over the origins, is also displayed in the table. From the table we see that DLM2 outperforms DLM1 for all forecast horizons, and especially for 18 months forecast. Not suprisingly, the forecast accuracy of DLM2 deteriorates as N increases but not dramatically so: the mean MAPE for N=6 is 3.5 %, compared to 4.9 % for N=18. The limited deterioration in forecast accuracy is not shared with DLM1, the ARIMA and Holt-Winters models, which both see substantial increases in forecasting errors as the forecast horizon increases. The mean MAPE of the Naïve model is stable across all forecast horizons. For short-term forecasts it is also one of the least accurate models. We see that the ETS model outperforms DLM2 for the shortest forecast horizon in terms of both RMSE and MAPE. For forecasts of 12 or 18 months the ETS model performs slightly better than DLM2 in terms of RMSE and slightly worse in terms of MAPE. Since squared errors gives a heavier penalty to large forecast errors than absolute errors, this means that DLM2 produces the smallest median forecast error and ETS the smallest mean forecast error. For forecast horizons of 12 and 18 months, the table shows that the errors of DLM2 has a small variance compared to other models. This suggests that the forecasting performance of this model generalises across the different training data sets.

Table 2: Mean (and standard deviation) of RMSE and MAPE

	Forecast horizon	ARIMA	DLM1	DLM2	ETS	Holt-Winters	Naïve	StructTS	TBATS
RMSE	N=6	0.4003	0.4315	0.3937	0.3781	0.4384	0.5693	0.4143	0.4131
		(0.095)	(0.1655)	(0.1087)	(0.0629)	(0.1822)	(0.193)	(0.1003)	(0.0654)
	N=12	0.5214	0.5841	0.4766	0.4676	0.609	0.5758	0.5396	0.5136
		(0.1434)	(0.2339)	(0.0982)	(0.1016)	(0.2443)	(0.124)	(0.1327)	(0.1403)
	N=18	0.6258	0.767	0.5297	0.5262	0.8085	0.6085	0.6702	0.6129
		(0.2435)	(0.3914)	(0.0884)	(0.1652)	(0.3599)	(0.1075)	(0.2245)	(0.231)
MAPE	N=6	0.0391	0.0426	0.0358	0.0362	0.0443	0.0546	0.0409	0.0411
		(0.0111)	(0.0171)	(0.0103)	(0.0079)	(0.0201)	(0.0196)	(0.0106)	(0.0084)
	N=12	0.0514	0.0577	0.0427	0.0446	0.0608	0.0535	0.0526	0.0514
		(0.0172)	(0.0262)	(0.0093)	(0.013)	(0.028)	(0.0115)	(0.0157)	(0.0156)
	N=18	0.0618	0.0752	0.0489	0.0509	0.08	0.0571	0.0648	0.0611
		(0.0269)	(0.041)	(0.0096)	(0.0188)	(0.0383)	(0.0122)	(0.0239)	(0.0249)

Figure 6 delves further into the variation in forecasting performance across origins, showing the MAPE per origin and model for the three forecast horizons. In the interval between December 2016 and December 2017, DLM2 is the best performing models for all forecast horizon, with the exception of origins right after March 2017 where the Naïve model perfoms best. We see that the ARIMA, Holt-Winters and the ETS models all perform better than the DLMs on the part of the time series where there is a clear trend in the training data that at least partially extends into the test data (until December 2016). As Table 2 showed, the higher forecast accuracy of the ETS model relative to the DLMs occurs primarily for a short forecast horizon. The relatively high MAPE for the DLMs prior to December 2016. Finally, the difference in performance between the two DLMs is, in general, minimal. However, the 6 month forecasts from medio 2017 onwards differ somewhat between the models, with DLM2 performing better than the DLM1.

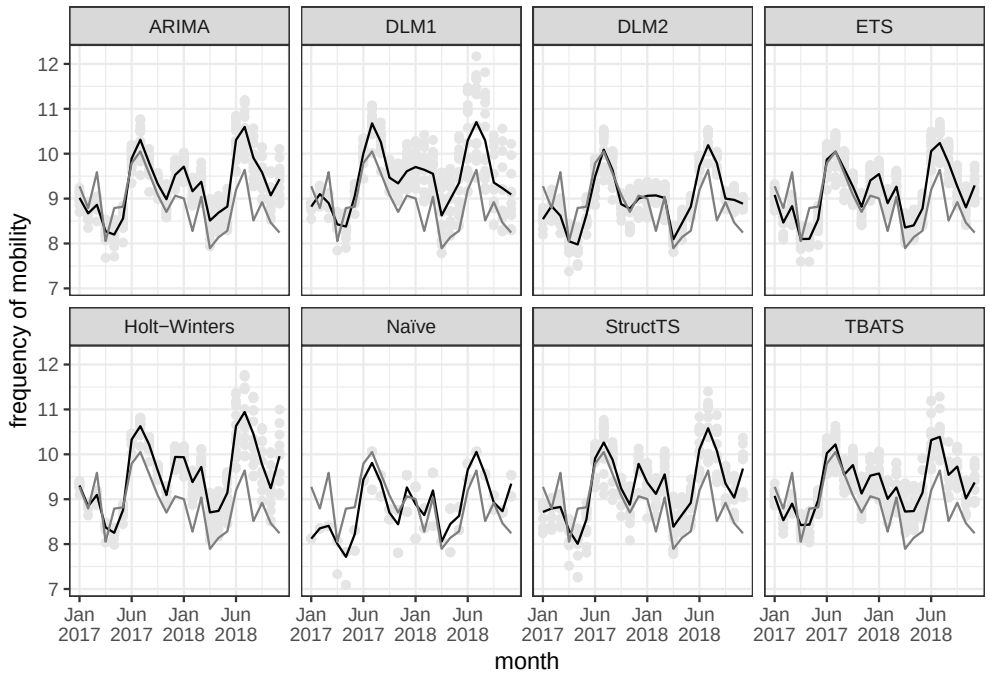
Figure 6: MAPE by origin and forecast horizon



4.2 Forecasted values from all origins

In addition to evaluating how forecasting performance vary between origins, we are also interested in the value of some specific forecasts. In particular, our initial interest in this exercise was motivated by a potential trend change. Therefore we zoom in on the forecasted values of the frequency of mobility around the changepoint. Figure 7 compares the forecasts of all models in the period January 2017 to December 2018. The grey dots represent forecast from different origins, the solid line represents the average value of the forecasts across all origins and the dotted line is the data. The figure shows clearly the divergence in forecasting performance between the DLMs and the ARIMA and Holt-Winters models. The Holt-Winters model systematically overpredicts the migration frequency already from June 2017, meaning that the forecasted frequency exceeds the data for all origins. For the ARIMA model this occurs from October 2017. In the case of the DLMs the last forecasted point with a value lower than that of the data occurs in May 2018.

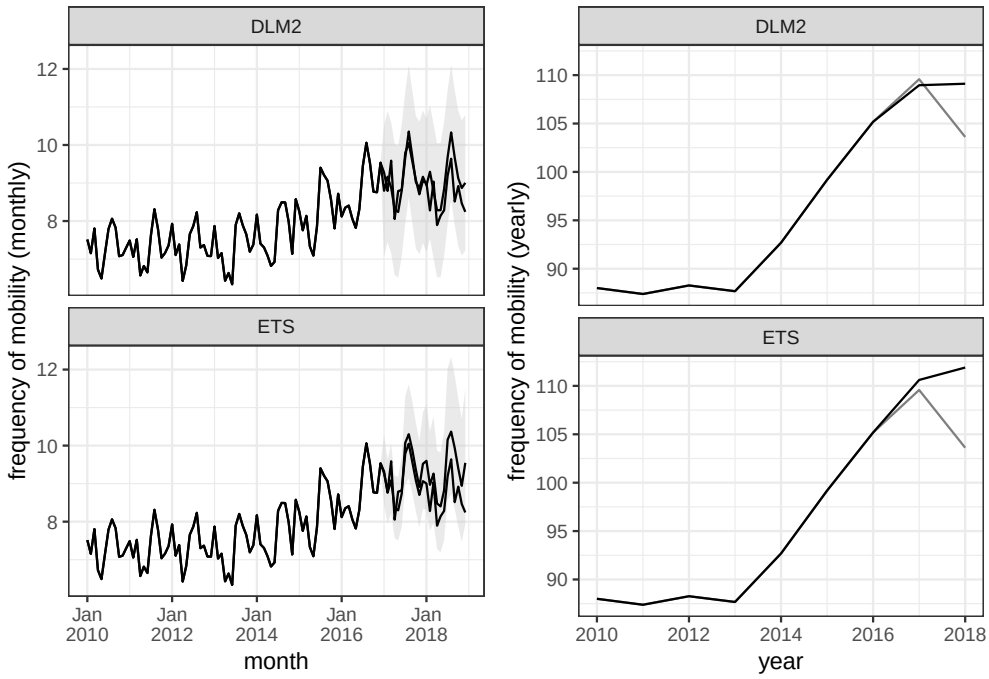
Figure 7: Forecast of frequency of mobility(light grey dots), forecasts averaged over origins (black line) and observed data (dark grey line)



4.3 Forecasting the annual frequency of mobility

As mentioned in the Introduction, the realisation that 2017 possibly represented a turning point in the cycle of the frequency of mobility triggered our initial interest in developing the state space model. Then it is relevant to ask what a model forecast can tell us about the development of the frequency of migration from 2017 onwards. In our case we were particularly interested in the development of the annual frequency in 2017 and 2018. The left panels in Figure 9 show a two-year forecast of the monthly and frequency of mobility with their 95 percent prediction intervals for the two best models from the previous subsection; namely ETS and DLM2. The automatic selection procedure results in an ETS(M, Ad, A): a model with multiplicative errors, damped additive trend and additive seasonal effects. The right panels show the yearly frequency calculated by aggregating the mean monthly forecast from the left panels. Note that the prediction intervals in the left panel are based on normality assumptions of the residuals of the model estimated on monthly frequency and can not be directly translated to annual frequency. We return to this issue below.

Figure 8: Two-year forecast (black line) of monthly (left) and annual (right) frequency of mobility and test data (dark grey line) and 95 percent prediction interval

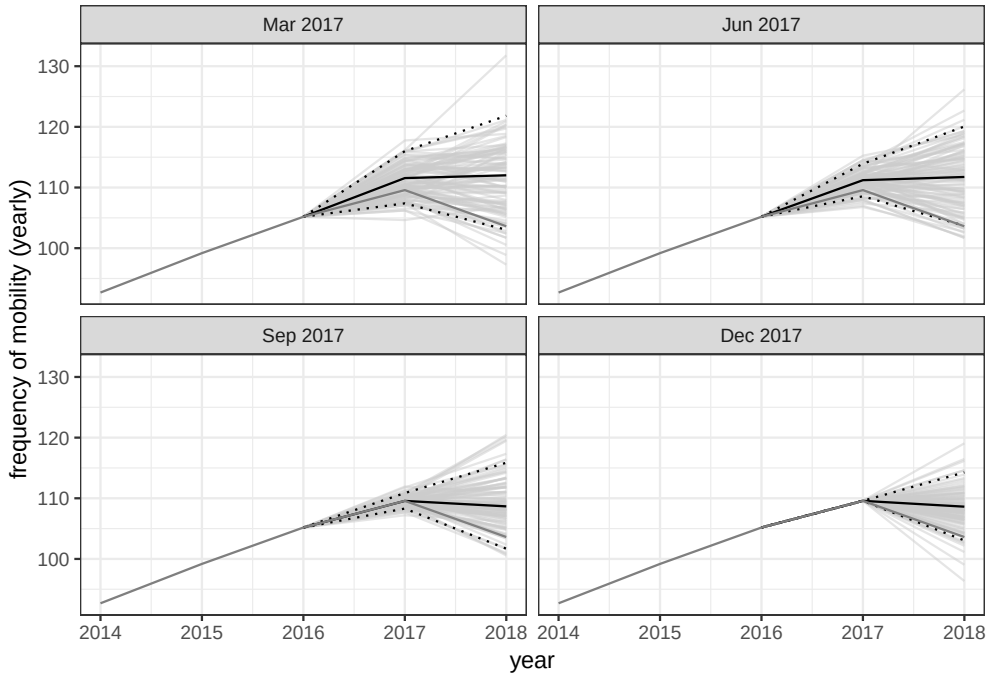


At first glance, the point forecast of the monthly frequency (left part of the figure) is quite similar across models, although the prediction interval of ETS is smaller. Both models forecast 2017 accurately, and they both systematically overestimate the test data in 2018. As the right left panels of Figure 9 shows, the annual frequency of mobility actually declines between 2017 and 2018, and none of the models were able to capture this sudden decline. However, we see ETS is much more "positive" about (i.e., overestimates more substantially) the annual frequency than DLM2. As we discussed earlier, both of these models can allow a short-run forecasts to diverge from a long-run trend growth. The figure thus suggests that the models produce quite similar short-run forecasts, but they differ in terms of their expected long-run growth. As a side note, the Naïve model would in fact predict the yearly frequency of 2018 very well.

Could additional monthly observations in 2017 help improve the accuracy of forecast of annual frequency in 2018? And how soon could we know that the top of the cycle had been reached? We can investigate these question using an approach similar to the evaluation on a rolling origin. For example, using the origin March 2017 we can generate forecasts for the rest of the months in 2017. By aggregating the observed (January to March) and

forecasted values (April to December) we can calculate the annual frequency of mobility in 2017. In order to generate forecasts, we utilise the fact that we can simulate from the model DLM2. This also allows us to calculate a Monte Carlo approximation of the prediction interval of the yearly frequency. Figure 9 shows the results from this exercise for four origins (March, June, September and December 2017), using 3000 simulated forecasts (only the first 100 are shown in the figure).

Figure 9: Two-year Monte Carlo forecast (light grey lines: forecasts; black line: average forecast value; dotted lines: boundaries of a 95 % Monte Carlo prediction interval) of yearly frequency of mobility and test data (dark grey line) for four different origins



Since we have identified March 2017 as the changepoint of the cycle, it is logical that this origin also results in the worst mean forecast (black line in Figure 9). Moving forward to the origin June 2017, we can see that the mean forecasts of 2017 and 2018 are closer to the observations and. From September onwards we see that the model clearly has picked up the development from after the changepoint, with the annual frequency declining. The 95 % prediction intervals include the observations for all origins while the intervals, logically, become progressively smaller as origins move forward in time.

5 Conclusion

In this paper we have developed two slightly different state space models for short-to medium term univariate forecast of monthly frequency of mobility in the Netherlands. The frequency of mobility is defined as the number of internally migrated persons countrywide (within- and between-municipality) per month divided by the population. This is an important input into the cohort-component model used for our regional population projections; the need for this modelling exercise arose as we expected that the end of our yearly series represented or was close to the top of a cycle.

The results presented in the previous section show that one of the models quite accurately forecasts both monthly and yearly frequency of mobility until end 2017, based on data up to end 2016. More generally, the model performed comparably or even better than a number of popular univariate forecasting in a time interval between end of 2016 to end of 2017. One of the most remarkable feature about the forecasts of the DLM was the relatively limited deterioration in forecast accuracy as the forecast horizon increased. We also showed how additional monthly data, released by Statistics Netherlands every month, allowed us to identify the changepoint of the cycle in the middle of a year.

Comparing the two DLMs, we saw that the specification of the local linear growth component played a crucial role. DLM2, which is in effect a linear trend model, ensured a certain conservatism for forecast horizons above 12 months (local trend model). The only other model mimicking this behaviour was the ETS model, itself also a state space model. An easy critique of these types of models is that a model that predicts that 'what goes up must eventually come down' is of limited value. Yet such criticism misses the point that the local linear growth component in effect incorporates not only local, but to some extent global, characteristics of the time series.

Although the time series used in this paper was an aggregate measure of mobility, it is of course well known that migration follows highly age-specific patterns (see e.g., Matthews and Parker 2013; Raymer, Willekens, and Rogers 2019). Disaggregating the time series into age groups (dynamic hierarchical models) may improve the forecast accuracy, and could be interesting for future research.

6 Acknowledgements

We thank Andries de Jong, Anet Weterings and Edwin Buitelaar (all senior researchers at PBL) for helpful comments and suggestions on the manuscript.

References

Beer, Joop De. 1993. Forecast intervals of net migration: The case of the Netherlands. *Journal of Forecasting* 12 (7): 585–599.

Bergmeir, Christoph, and José M Benıtez. 2012. On the use of cross-validation for time series predictor evaluation. *Information Sciences* 191:192–213.

Bergmeir, Christoph, Rob J. Hyndman, and Bonsoo Koo. 2018. A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis* 120:70–83. ISSN: 0167-9473.

Canova, Fabio. 1998. Detrending and business cycle facts. *Journal of monetary economics* 41 (3): 475–512.

de Jong, Andries, M Alders, P Feijten, Petra Visser, Ingeborg Deerenberg, M Van Huis, and Dick Leering. 2005. Achtergronden en veronderstellingen bij het model PEARL. Naar een nieuwe regionale bevolkings-en allochtonenprognose. *Den Haag: Ruimtelijk Planbureau/Centraal Bureau voor de Statistiek*.

De Livera, Alysha M, Rob J Hyndman, and Ralph D Snyder. 2011. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American statistical association* 106 (496): 1513–1527.

Durbin, James, and Siem Jan Koopman. 2012. *Time series analysis by state space methods*. Oxford university press.

Hamilton, James D. 2018. Why you should never use the Hodrick-Prescott filter. *Review of Economics and Statistics* 100 (5): 831–843.

Harvey, Andrew C. 1990. *Forecasting, structural time series models and the Kalman filter*. Cambridge university press.

Holt, Charles C. 2004. Forecasting seasonals and trends by exponentially weighted moving averages. *International journal of forecasting* 20 (1): 5–10.

Husby, Trond, Anet Weterings, and Jolien de Groot. 2019. *Trek van en naar de stad. veranderingen in verhuispatronen, 1996-2018*. Den Haag: Planbureau voor de Leefomgeving. `https://themasites.pbl.nl/trek-van-en-naar-de-stad/`.

Hyndman, Rob J, and George Athanasopoulos. 2018. *Forecasting: principles and practice*. OTexts.

Hyndman, Rob J, and Yeasmin Khandakar. 2008. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software* 26 (3): 1–22.

Hyndman, Rob J, Anne B Koehler, Ralph D Snyder, and Simone Grose. 2002. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of forecasting* 18 (3): 439–454.

Kalman, Rudolf Emil. 1960. Contributions to the theory of optimal control. *Boletin de la Sociedad Matematica Mexicana* 5 (2): 102–119.

Kaplan, Greg, and Sam Schulhofer-Wohl. 2017. Understanding the long-run decline in interstate migration. *International Economic Review* 58 (1): 57–94.

Makridakis, Spyros, Rob J Hyndman, and Fotios Petropoulos. 2019. Forecasting in social settings: the state of the art. *International Journal of Forecasting*.

Matthews, Stephen A, and Daniel M Parker. 2013. Progress in spatial demography. *Demographic Research* 28:271–312.

Monahan, John F. 1984. A note on enforcing stationarity in autoregressive-moving average models. *Biometrika* 71 (2): 403–404.

Mulder, Clara H. 2018. Putting family centre stage: Ties to nonresident family, internal migration, and immobility. *Demographic Research* 39:1151–1180.

Petris, Giovanni, Sonia Petrone, and Patrizia Campagnoli. 2009. *Dynamic linear models with R*. Springer.

Raymer, James, Frans Willekens, and Andrei Rogers. 2019. Spatial demography: A unifying core and agenda for further research. *Population, Space and Place* 25 (4): e2179.

Smith, Stanley K. 1997. Further thoughts on simplicity and complexity in population projection models. *International journal of forecasting* 13 (4): 557–565.

Smith, Stanley K, Jeff Tayman, and David A Swanson. 2013. *A practitioner's guide to state and local population projections*. Springer.

Tashman, Leonard J. 2000. Out-of-sample tests of forecasting accuracy: an analysis and review. The M3- Competition, *International Journal of Forecasting* 16 (4): 437–450. ISSN: 0169-2070.

te Riele, Saskia, Corina Huisman, Lenny Stoeldraijer, Andries de Jong, Coen van Duin, and Trond Husby. 2019. PBL/CBS Regionale bevolkings- en-huishoudensprognose 2019–2050: Belangrijkste uitkomsten. *Statistische Trends*.

Winters, Peter R. 1960. Forecasting sales by exponentially weighted moving averages. *Management science* 6 (3): 324–342.

Young, Peter Colin, Cho Nam Ng, Kevin Lane, and David Parker. 1991. Recursive forecasting, smoothing and seasonal adjustment of non-stationary environmental data. *Journal of Forecasting* 10 (1-2): 57–89.

Zietz, Joachim, and Anca Traian. 2014. When was the US housing downturn predictable? A comparison of univariate forecasting methods. *The Quarterly Review of Economics and Finance* 54 (2): 271–281.

Appendix A

In addition to the cycle, there are strong seasonal patterns that seem to vary over time. The seasonal subseries plot in Figure 10 shows the movement per year and average per month (lower panel), and the horizontal lines in the figure indicating the means for each month. This plot enables us to see the underlying seasonal pattern clearly, and it also shows the changes in seasonality over time. The figure shows that the highest frequency is, on average, in August and July while the lowest is in April. We also see there is quite some variation between the years: the differences between the months were more pronounced in the early part of the time series than in later years. The lower panel suggests a sinusoidal pattern with two peaks within one year - one peak in the summer and one at the end of the year.

Finally, we check whether there is a linear relationship between lagged variables of the time series (autocorrelation). Figure 11 reveals a large and positive autocorrelation for small lags, since observations nearby in time tend to be similar in size. We also see that strong autocorrelation in lags that are multiples of the seasonal frequency (12, 24, and 36), which is due to the seasonality discussed above. The slow decline is related to the trend-cycle while the 'scalloped' pattern is related to the seasonality. In terms of lag selection for an ARIMA model, the significant spikes at the first and second lag in the partial autocorrelation plot in the lower figure suggests that we should include at least two autoregressive terms (Hyndman and Athanasopoulos 2018).

Figure 10: Seasonal subseries plot of the frequency of mobility

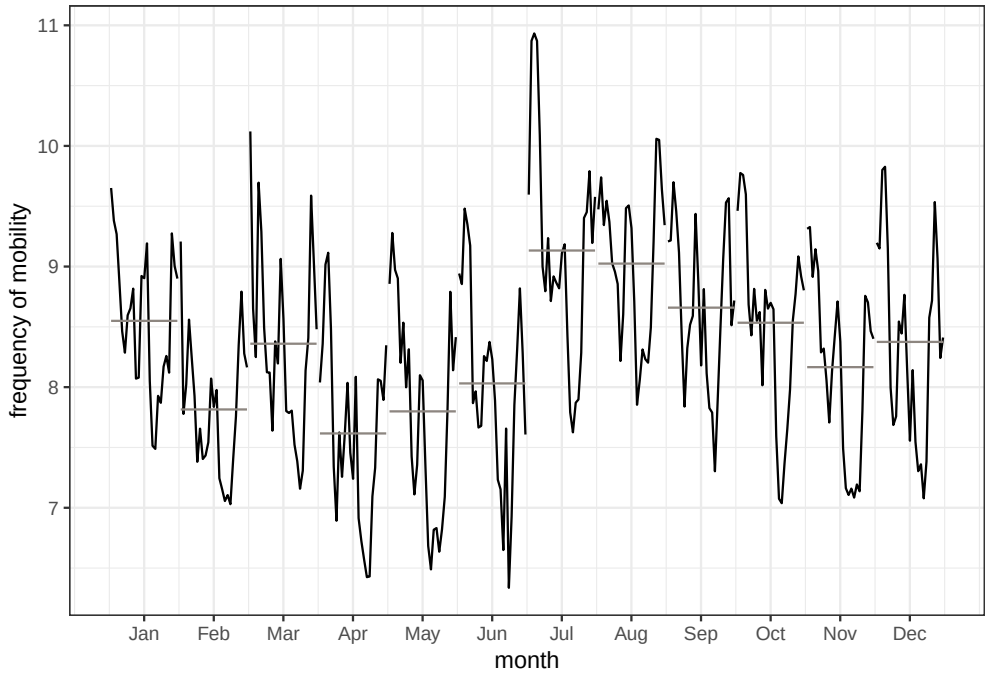
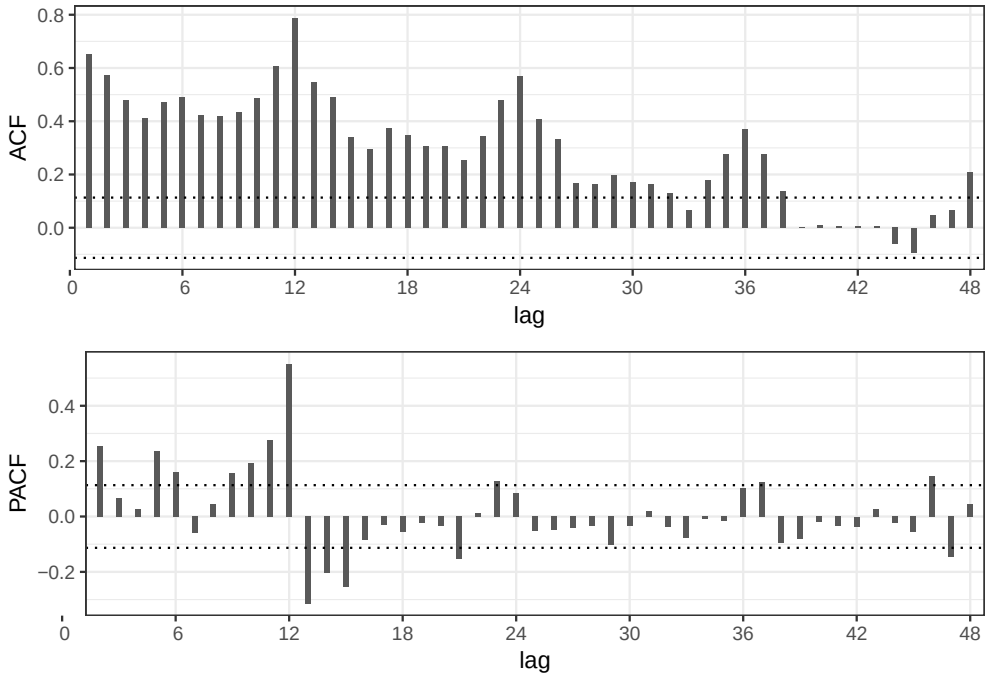


Figure 11: Autocorrelation function of the frequency of mobility



Appendix B

Figure 12: Autocorrelation function of the standardised residuals

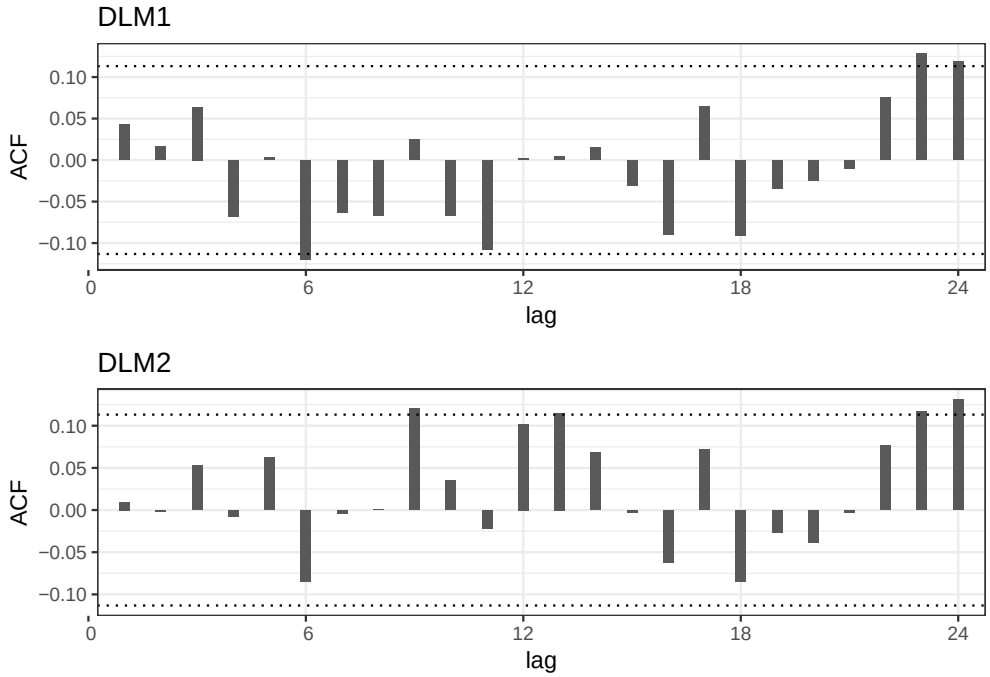


Figure 13: Normal probability plot of standardized one-step-ahead forecast errors of DLM2

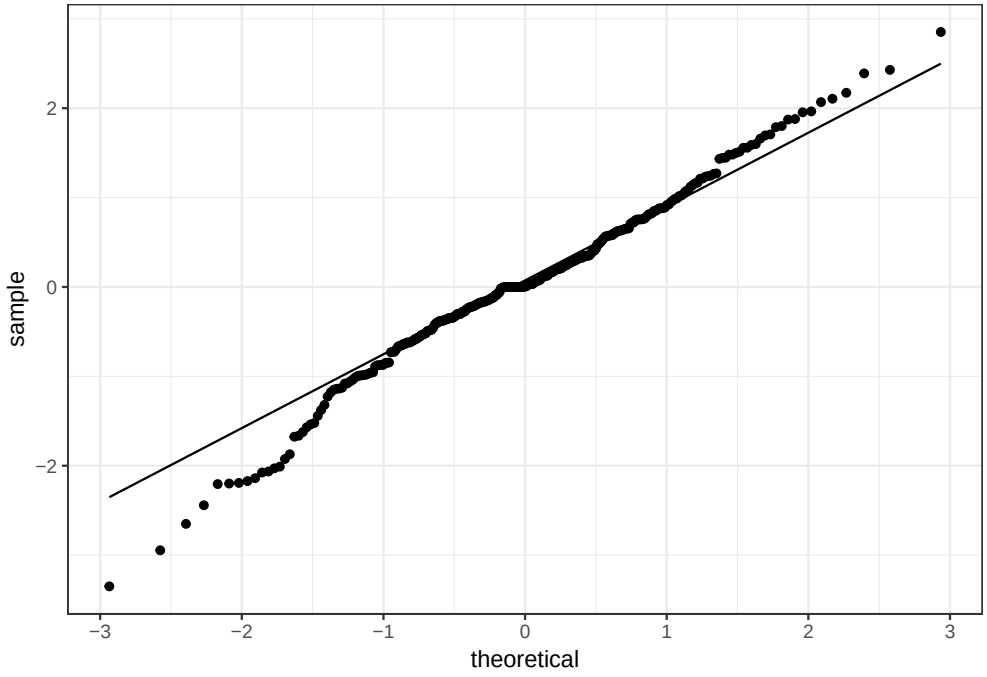


Table 3: Mean and standard deviation of the estimated parameters across all origins

Parameter	DLM1 Mean	DLM1 SD	DLM2 Mean	DLM2 SD
σ_v^2	1.523e-08	5.459e-13	1.523e-08	8.178e-14
σ_μ^2	0	0	0.000152	4.744e-06
σ_β^2	1.828e-06	2.717e-07	0	0
$\sigma_{S_2}^2$	7.899e-06	2.844e-07	7.665e-06	2.572e-07
σ_u^2	0.001342	1.913e-05	0.001028	1.236e-05
ϕ_1	-0.6689	0.01432	-0.7349	0.01313
ϕ_2	0.07194	0.008122	-0.03206	0.005823
ϕ_7	-0.445	0.005193	-0.5018	0.00541
ϕ_{12}	0.4363	0.003109	0.4255	0.002562

Figure 14: Filtering state estimates of the slope (β_t)