

Forecasting internal migration with a state space model

Trond

1 Introduction

Long-term projections of internal migration are key inputs to regional-level demographic population projections. The long-term projections rely on the identification of a long-term trend in data, thus filtering out short-run fluctuations due to the business cycle. However, such detrending analyses are complicated if there is no clear trend in the data, if business cycles themselves stretch into the medium- to long run or if the data point represent the top or the bottom of the cycle (Canova 1998; Hamilton 2018). Therefore it is informative to analyse the short- to medium term dynamics of demographic components, forecasting along the business cycle. This paper develops a state space model for short-to medium term univariate forecast of monthly frequency of mobility in the Netherlands.

As is the case in most developed countries, the main driver of regional-level population change in the Netherlands is migration (external and domestic). Due to the close relation between migration decisions and labour- and housing market conditions, the volume of internal migration often moves along with the business cycle. However, recent research argues that the relationship between economic drivers and migration has changed (Kaplan and Schulhofer-Wohl 2017). And other drivers than macroeconomic factors may be just as important in explaining migration - for example, family considerations are important determinants of migration decisions on an individual level, meaning migration flows could also be affected by changes to family composition and aging (see e.g. Mulder 2018). As such, there are many possible candidate determinants of internal migration and their influence on migration flows may change over time.

Even if one succeeds in finding these determinants, their practical use in regional population projections may be somewhat limited. 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. 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, in practice the growth path used as inputs in cohort-component models is most commonly based on extrapolation of long-term historic trends.

When working with the latest Dutch Regional Population Projections, published in September 2019 (te Riele et al. 2019), we were questioning whether the last data point in the time series of yearly frequency of mobility - registered at the end of 2016 - represented the top of a cycle. We knew, for example, that a recent change to the Dutch student finance system seems to have caused a decrease in mobility for an otherwise very mobile group.¹ As is common in the literature, key inputs for the cohort-component model used for the projections are extrapolations of the long-term trend of the components. Obviously, if the last data point used to determine the long-term trend represents the top of a cycle, our extrapolation could overestimate the true trend. In order to determine whether the last point of our yearly series of the migration frequency represented the top of a cycle, it was informative to carry out a medium-term forecast based on data which was both more recent and of higher frequency than the yearly series.

Forecasting of time series often involves decomposing the series into the more basic elements such as trend, cycle and seasonality. Univariate forecasting models make use of basic time series patterns to form a forecast (Zietz and Traian 2014). One popular type of model used for this purpose are Dynamic Linear Models (DLM), which is a type of State Space models (Petrís, Petrone, and Campagnoli 2009; Durbin and Koopman 2012). As the name suggests, DLMs are linear models with a Gaussian error structure, where the relevant inferences are carried out using the Kalman filter. Unknown parameters of the model are estimated by either maximum likelihood or Bayesian techniques. One advantage of this model type is its modularity; time series components can be added to or removed from the model whenever deemed necessary. Another advantage is that state space models are stochastic, meaning one can derive forecast distributions, either analytically or by simulation.

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

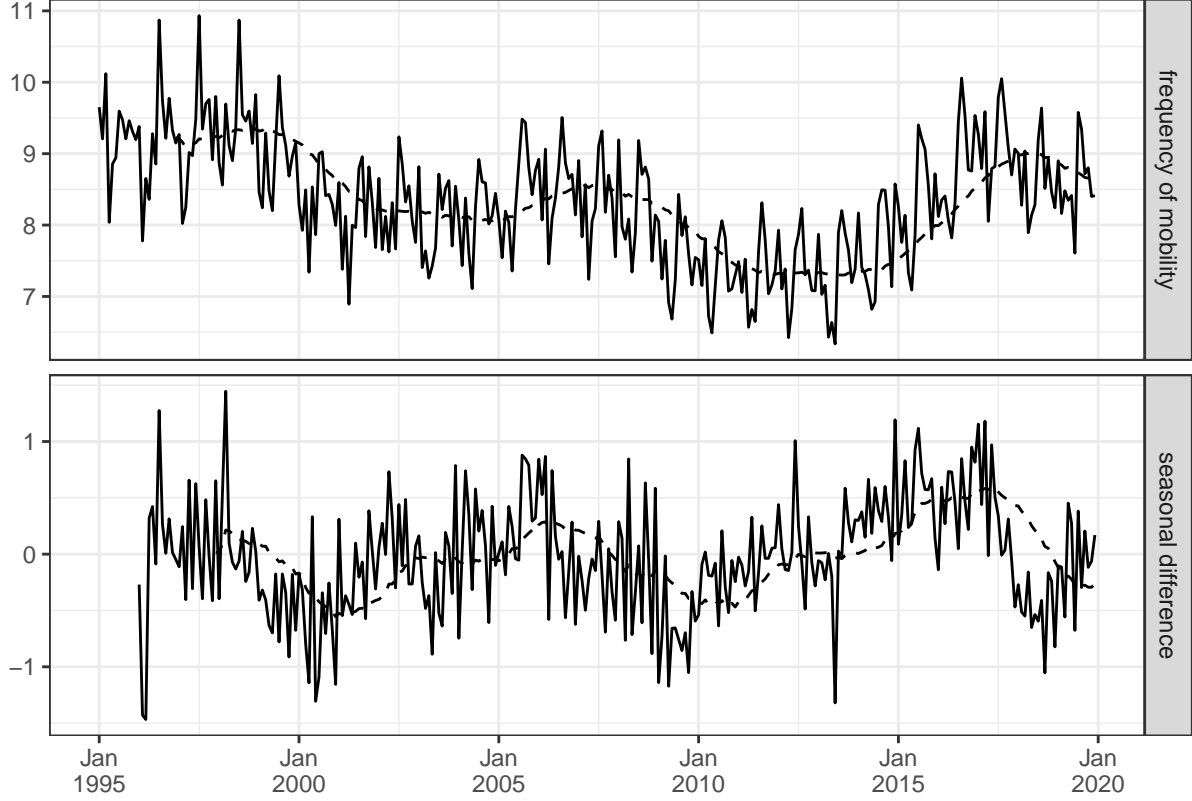


Figure 1: Frequency of mobility (upper panel) and seasonal differences (lower panel) and two-year moving averages (dotted line)

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. This represents a measure of the intensity of internal migration in the Netherlands, and gives us a rough picture of migration also on a regional level. We compare the forecasting performance on our model with that of four other popular models for univariate time-series forecasting. These models 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.

In the paper I firstly show the time series data and discuss its patterns. Next I present the model, discuss its elements and the estimated parameters. Thereafter I evaluate the out-of-sample forecasting performance of the DLM, comparing it to that of four other models. Finally I discuss what the models tell us about the development of the frequency of mobility from 2017 onwards.

2 Data

The frequency of mobility y_t in month t is in this paper defined as the sum of inter- en intramunicipality moves between the first and the last day of month t , divided by the population on the last day of t . 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. Figure 1 shows the time series from January 1995 to December 2019 as well as the seasonal differences $y_t' = y_t - y_{t-12}$ as well as the monthly differences $y_t^* = y_t - y_{t-1}$. The dotted line in the figure shows two-year moving averages of both the frequency of mobility and the seasonal differences.

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 middle 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. This shows that our suspicions of a peak in mobility frequency at the end of 2016 were not far off the mark. As suggested by the lower panel, the monthly differences of the frequency of mobility approximates a random walk. The visual inspection of this figure reveals a fairly unpredictable long-term pattern, with jumps hovering around zero. The moving averages suggest indeed that the medium-term average monthly change of the series is zero.

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 these elements are addressed in the state space model.

3 Model description

The “state” in state space models is an unobserved dynamic process θ_t for which only a noisy measurement y_t can be observed. State space models are modular in the sense that the state vector θ_t is constructed by adding components deemed necessary to characterise the time series. As mentioned in the Introduction, DLMS are a special case of state space models with Gaussian and independent errors. Forecasting and inference is carried out with the recursive Kalman filter algorithm (Kalman and others 1960). Residuals are calculated by comparing the one-step ahead predictions from the Kalman filter with the observed values. There are many ways of describing state space models; 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 , where y_t are the observed values of underlying (unobserved) vector of states θ_t . y_t are conditionally independent given the state θ_t . The state represents 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:

$$\begin{aligned} y_t &= F\theta_t + v_t & v_t &\sim N(0, V_t) \\ \theta_t &= G\theta_{t-1} + w_t & w_t &\sim N(0, W_t) \\ \theta_0 &\sim N(m_0, C_0) \end{aligned}$$

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 .

As mentioned, elements of the state vector θ_t should be chosen such that they reflect the characteristics of the time series in question. The first thing we noted in the previous section was that the time series was clearly non-stationary with no global trend. However, we saw that periods of growth were followed by periods of decline, meaning there were piecewise or local 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

²<https://www.cbs.nl/nl-nl/nieuws/2018/26/minder-mensen-verhuisd-in-eerste-kwartaal-2018>

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:

$$\begin{aligned} y_t &= \mu_t + v_t & v_t &\sim N(0, \sigma_v^2) \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + w_{t,1} & w_{t,1} &\sim N(0, \sigma_\mu^2) \\ \beta_t &= \beta_{t-1} + w_{t,2} & w_{t,2} &\sim N(0, \sigma_\beta^2) \end{aligned}$$

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

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

Seasonality is dealt with by including a trigonometric seasonal components to the model with a fixed periodicity of 12 months. One big 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 through time. Seasonality is incorporated in the DLM model by extending the state equation with an underlying periodic process g_t , defined as the sum of j harmonics. For the j th harmonic, we can write the evolution of the seasonal effects as:

$$\begin{aligned} S_{j,t+1} &= S_{j,t} \cos \omega_j + S_{j,t}^* \sin \omega_j \\ S_{j,t+1}^* &= -S_{j,t} \sin \omega_j + S_{j,t}^* \cos \omega_j \end{aligned}$$

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

$$\theta_t = \begin{bmatrix} S_{1,t} \\ S_{1,t}^* \\ S_{2,t} \\ S_{2,t}^* \end{bmatrix} \quad G = \begin{bmatrix} \cos \omega_1 & \sin \omega_1 & 0 & 0 \\ -\sin \omega_1 & \cos \omega_1 & 0 & 0 \\ 0 & 0 & \cos \omega_2 & \sin \omega_2 \\ 0 & 0 & -\sin \omega_2 & \cos \omega_2 \end{bmatrix} \quad F = \begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix} \quad W = \begin{bmatrix} \sigma_{S_1}^2 & 0 & 0 & 0 \\ 0 & \sigma_{S_1^*}^2 & 0 & 0 \\ 0 & 0 & \sigma_{S_2}^2 & 0 \\ 0 & 0 & 0 & \sigma_{S_2^*}^2 \end{bmatrix}$$

The two components discussed so far (trend and seasonal) remove much of the autocorrelation, however a Ljung-Box test confirmed that the residuals of a model consisting of only were still not white noise. One way of dealing with the residual autocorrelation is to include autoregressive elements into the state space model. On the basis of the partial autocorrelation function show in Appendix A, and further analysis with an AR(2) model we choose to include an AR(4) with two higher order lags - namely lags 7 and 12. Consequently, the autoregressive component of our state space model is an AR(4) model with ‘gaps’. This can be cast in state space form 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

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

Notice that the standard routine for creating autoregressive elements in the *dlm* package do not allow for ‘gaps’ in the included lags. In order to estimate such a model we need to implement system matrices of an AR(12) where the parameter of lags not included in the model are set to zero:

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

The resulting composite model (trend, seasonal and autoregressive) has a state vector θ_t of length 7: $\theta_t = (\mu_t, \beta_t, S_{1,t}, S_{1,t}^*, S_{2,t}, S_{2,t}^*, \mu_{1,t})'$. In order to apply the Kalman filter we first need to determine the unknown model parameters. Unknown parameters are the variances of the error terms, in effect hyperparameters, and the parameters for the autoregressive terms. We chose to estimate all of them using maximum likelihood (see Petris, Petrone, and Campagnoli 2009, ch. 4). As we have seen, the matrix W contains 9 potentially non-zero parameters: σ_v^2 , σ_μ^2 , $\sigma_{S_2}^2$, σ_u^2 , ϕ_1 , ϕ_2 , ϕ_7 , and ϕ_{12} . Note that the parameters governing the autoregressive component are estimated subject to stationarity restrictions (Monahan 1984).

4 Results

4.1 Estimated parameters

As we saw in the previous section, the trend component contains two error terms with their own variances - one for the level and one for the slope. Initial runs with the model discussed so far revealed that two special cases fit the data better than the general formulation. 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 (REF). In this specification, all noise is captured by the slope, meaning that the local linear trend 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 growth of the time series rather than the local growth. As we saw in Figure 1, the slope representing long-run growth (moving averages of the monthly differences) is almost completely flat. However, 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 . If the time series diverges from the long-run growth in this interval, this will also be picked up by the forecasts. 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.

4.2 Filtering and smoothing estimates of the frequency of mobility

With estimates of the unknown parameters we use the Kalman filter to obtain the filtered distribution of θ_t conditional on the observed series up to t , $\theta_t|y_1, y_2, \dots, y_t$. 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 Y_{t+1} . Similarly, an n -step ahead forecast Y_{t+n} is based on calculation of the n -step ahead state vector θ_{t+n} . We can write the forecast function as $f_t = E(Y_{t+n}|y_1, y_2, \dots, y_t)$. In addition we estimate a smoothing distribution of the components, representing the past values of the states given all observed values $\theta_t|y_1, y_2, \dots, y_s$ where $s \geq t$.

The one-step ahead predictions from the Kalman filter, or filtered values, from both of the models are displayed in Figure 2, together with Root Mean Square Error (RMSE) calculated on basis of the residuals (one-step-ahead forecast errors). The adaptive nature of the Kalman algorithm causes the filtered values in

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
$\sigma_{S_2}^2$	7.723e-06	7.723e-06
σ_u^2	0.001284	0.001022
ϕ_1	-0.6428	-0.6957
ϕ_2	0.03474	-0.05385
ϕ_7	-0.4429	-0.4899
ϕ_{12}	0.451	0.4359

the first periods to veer quite far off the observed values (not shown in the figure). We therefore discard the three first years when calculating the RMSE. It is clear from the figure that DLM2 fits the data somewhat better than the DLM1 model, though the difference is not dramatic. A residual check showed that the normality assumption of dynamic linear models is justified for these models (see Appendix B for a residual plot of DLM2).

Figure 3 shows the one-step ahead predictions 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 (Petrís, Petrone, and Campagnoli 2009, ch. 3). From this plot we clearly see that DLM1 systematically predicts a higher frequency of mobility than DLM2 from 2014 to about 2017. Although barely visible in the figure, the prediction interval of DLM1 is slightly smaller than that of DLM2.

The smoothing estimates for the trend-cycle component and for the seasonal effects are shown in Figure 4. From the left panels we can pin-point 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. The right panels show how the seasonal effects vary over time; at the start of the time series the within-year cycle exhibits 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.

5 Discussion

5.1 Cross validation: evaluation over a rolling origin

In this section we evaluated 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). For cross-sectional data such 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 using k-fold cross validation. Due to the serial autocorrelation and potential non-stationarity in the data, it is common to choose a test set that does not contain observations occurring prior to the observations in the training data (Bergmeir and Benítez 2012). Recent literature suggests that standard cross-validation can be applied to certain time series models (Bergmeir, Hyndman, and Koo 2018) - however, due the Markovian structure of the Kalman filtering algorithm, we suspect that the technique would not work for the state space models discussed in the previous section. Therefore we follow the conventional approach for evaluating time series accuracy; namely evaluation over a *rolling forecasting origin* (Tashman 2000), where we succesively extend the training data and produce forecasts of the test data. The last point of the training data in each iteration is referred to

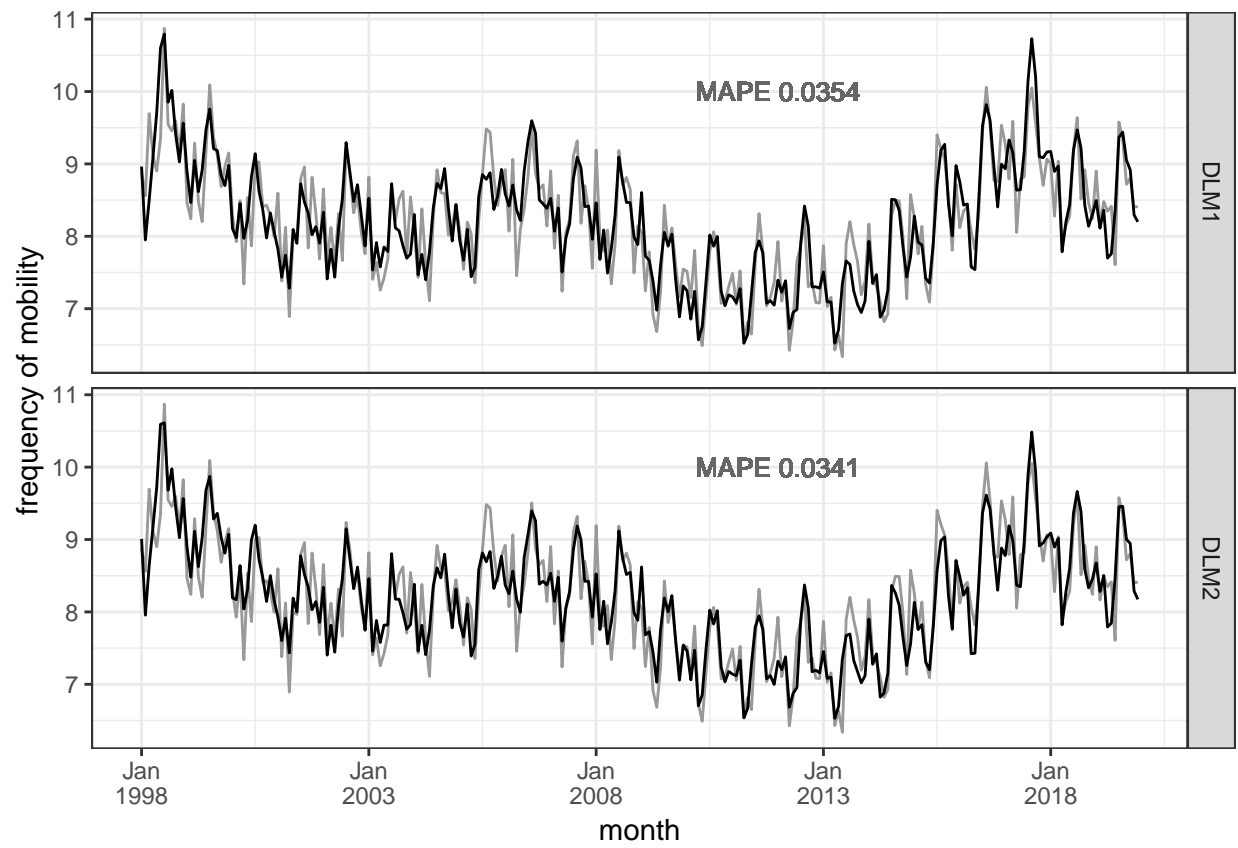


Figure 2: One-step-ahead predictions (solid line), observed data (dotted line) and RMSE (calculated from January 1998)

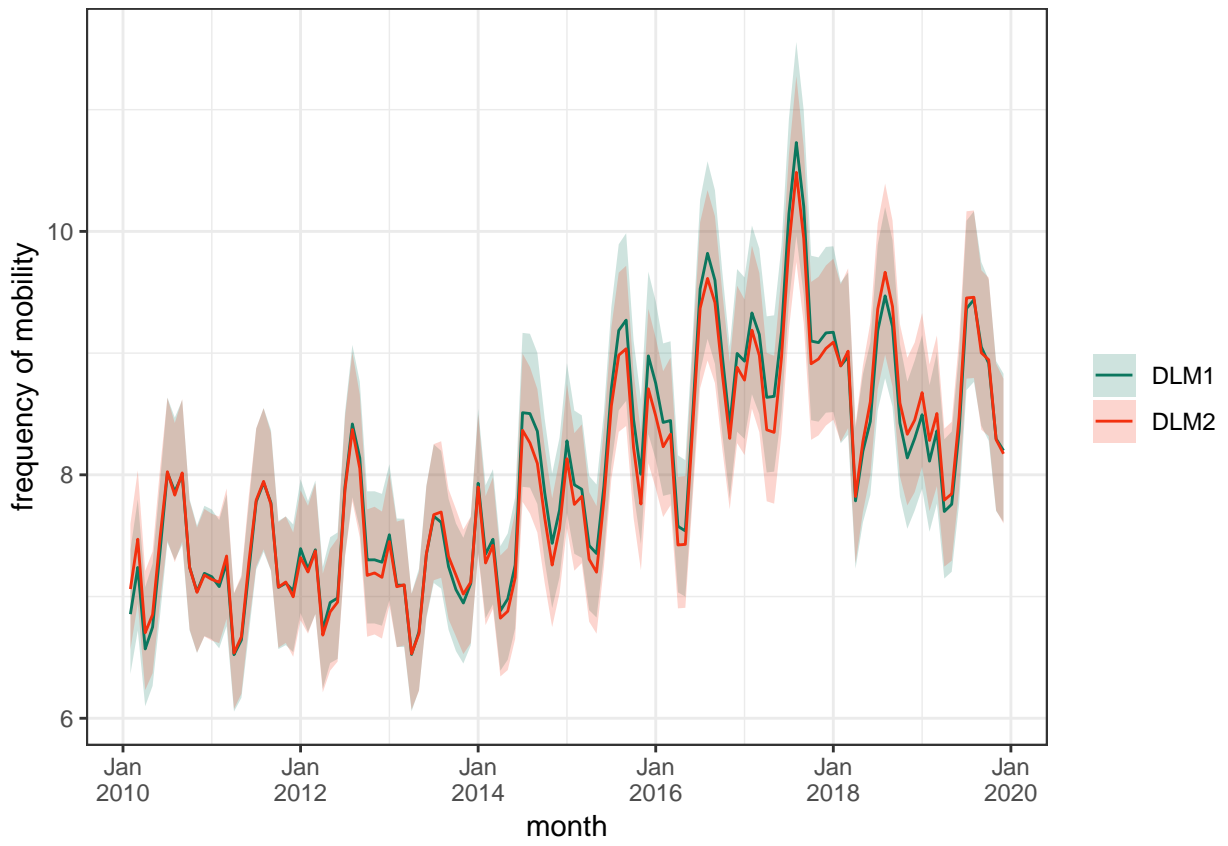


Figure 3: Filtering estimate and 95 percent prediction intervals

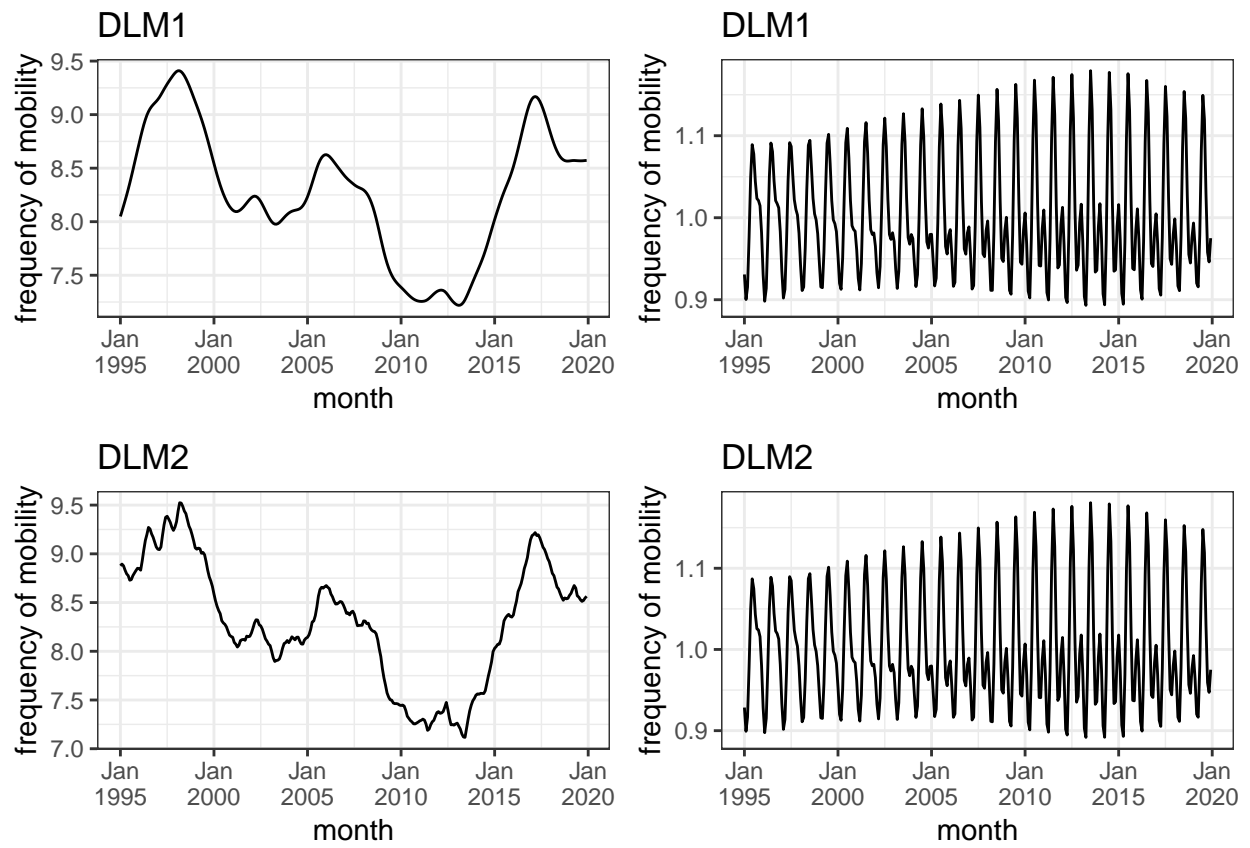


Figure 4: Smoothing estimate of the trend-cycle and seasonality

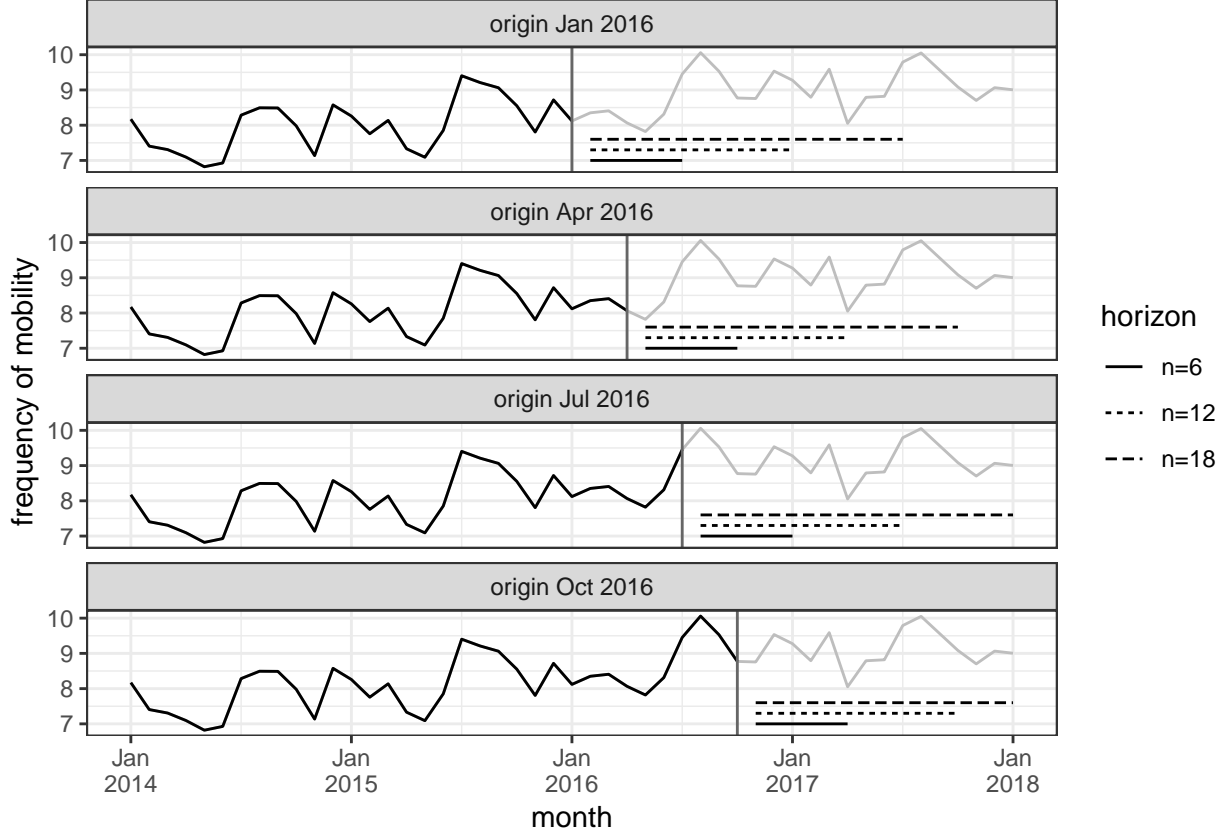


Figure 5: Illustration of rolling forecasting origin: training data (solid black line), origin (vertical dark grey line), forecast horizon (dotted horizontal lines) and test data (grey line)

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 $t = 1, \dots, T$ and forecasts are generated for time periods $T + 1, T + 2, \dots, T + N$. 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). The model performance is then evaluated by averaging the prediction errors over different forecast horizon (dotted horizontal lines).

We evaluate forecasting performance on three time horizons - $N = 6$, $N = 12$ and $N = 18$ - meaning 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 forecasting origins: $T =$ March 2016, February 2016, \dots , June 2017.

In order to ensure independence between the test and training sample, we estimate the models on the relevant training data of each origin forecast. As such, 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 three 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 is useful with data with no clear trends. Since the model assumes no trend, we expect this model to perform particularly well around the

changepoint. 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 a trend. 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, allowing for a much more general models of exponential smoothing methods. We will refer to this model type as ETS (Error, Trend, Seasonal). An advantage of casting 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 (R. J. 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).

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 (R. J. 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 all origins for all models. This investigating the distribution of accuracy across “folds” (in this case forecasting origins). The table shows both RMSE and mean absolute percentage error (MAPE), averaged over origins, of each of forecasting horizon (6 months, 12 months, 18 months). In order to assess whether the accuracy varies between origins the table also shows the standard deviation of RMSE and MAPE. From the table we see that both RMSE and MAPE of the DLM2 are quite similar across all forecast horizons, however with a slight advantage of DLM2. Not surprisingly, accuracy deteriorates as *N* increases, although not dramatically so: in the case of DLM2, the mean RMSE for *N*=6 is 0.39, compared to 0.53 for *N*=18. The limited deterioration in forecast accuracy is not shared with the ARIMA and Holt-Winters models, which both see substantial increases in forecasting errors as the forecast horizon increases. The mean RMSEs of the Naïve model is remarkably stable across all forecast horizons, and it is, in general, the least accurate model. 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.

Figure 6 delves further the variation in forecasting performance across origins, showing the MAPE per origin and model for the three forecast horizons. In all forecast horizons the two DLMS are the best performing models in the interval between December 2016 and December 2017, with the exception of origins right after March 2017 where the Naïve model performs 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 showed how the forecasting performance varied between origins; it suggested that the DLMS performed poorly, relative to other models, where a trend carried over from the training data to the test data and good relative to other models otherwise. Of course 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

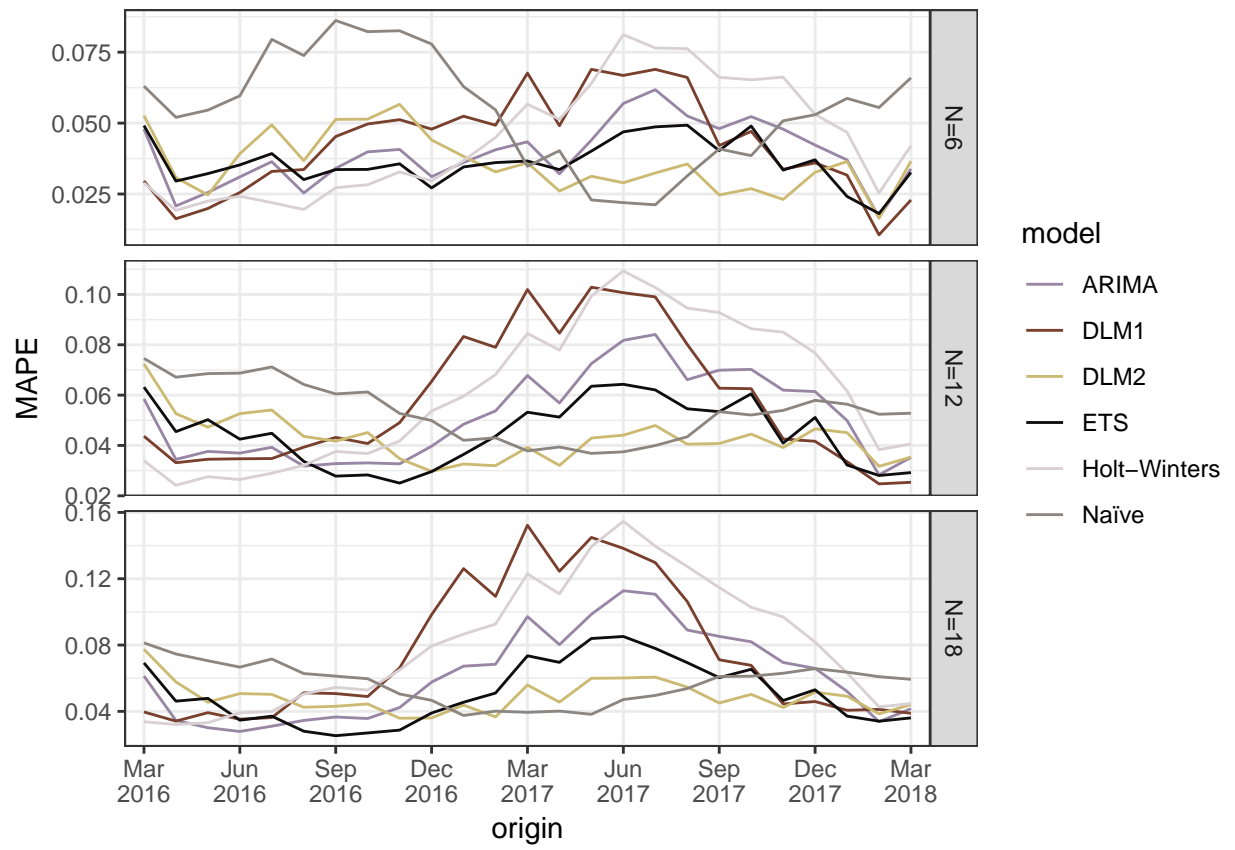


Figure 6: MAPE by origin and forecast horizon

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

	Forecast horizon	ARIMA	DLM1	DLM2	ETS	Holt-Winters	Naïve
RMSE	N=6	0.4003	0.4315	0.3937	0.3781	0.4384	0.5693
		(0.095)	(0.1655)	(0.1087)	(0.0629)	(0.1822)	(0.193)
	N=12	0.5214	0.5841	0.4766	0.4676	0.609	0.5758
		(0.1434)	(0.2339)	(0.0982)	(0.1016)	(0.2443)	(0.124)
	N=18	0.6258	0.767	0.5297	0.5262	0.8085	0.6085
		(0.2435)	(0.3914)	(0.0884)	(0.1652)	(0.3599)	(0.1075)
MAPE	N=6	0.0391	0.0426	0.0358	0.0362	0.0443	0.0546
		(0.0111)	(0.0171)	(0.0103)	(0.0079)	(0.0201)	(0.0196)
	N=12	0.0514	0.0577	0.0427	0.0446	0.0608	0.0535
		(0.0172)	(0.0262)	(0.0093)	(0.013)	(0.028)	(0.0115)
	N=18	0.0618	0.0752	0.0489	0.0509	0.08	0.0571
		(0.0269)	(0.041)	(0.0096)	(0.0188)	(0.0383)	(0.0122)

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.

5.2 Forecasting migration frequency from January 2017 onwards

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 the second state space model 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 yearly frequency the coming couple of years. Figure 8 shows a two-year forecast of both the monthly and yearly frequency of mobility with their 95 percent prediction intervals, where the last period of the training data is December 2016. For simplicity, the yearly frequency is calculated by aggregating the monthly frequencies within that year. As a comparison, we also show results the same forecast using the ETS model discussed in the previous section. The automatic selection procedure results in an ETS(M, Ad, A): a model with multiplicative errors, damped additive trend and additive seasonal effects.

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 8 shows, the yearly 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 yearly 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. But a clock without battery also shows the right time twice a day.

6 Conclusion

In this paper we have developed a state space model 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

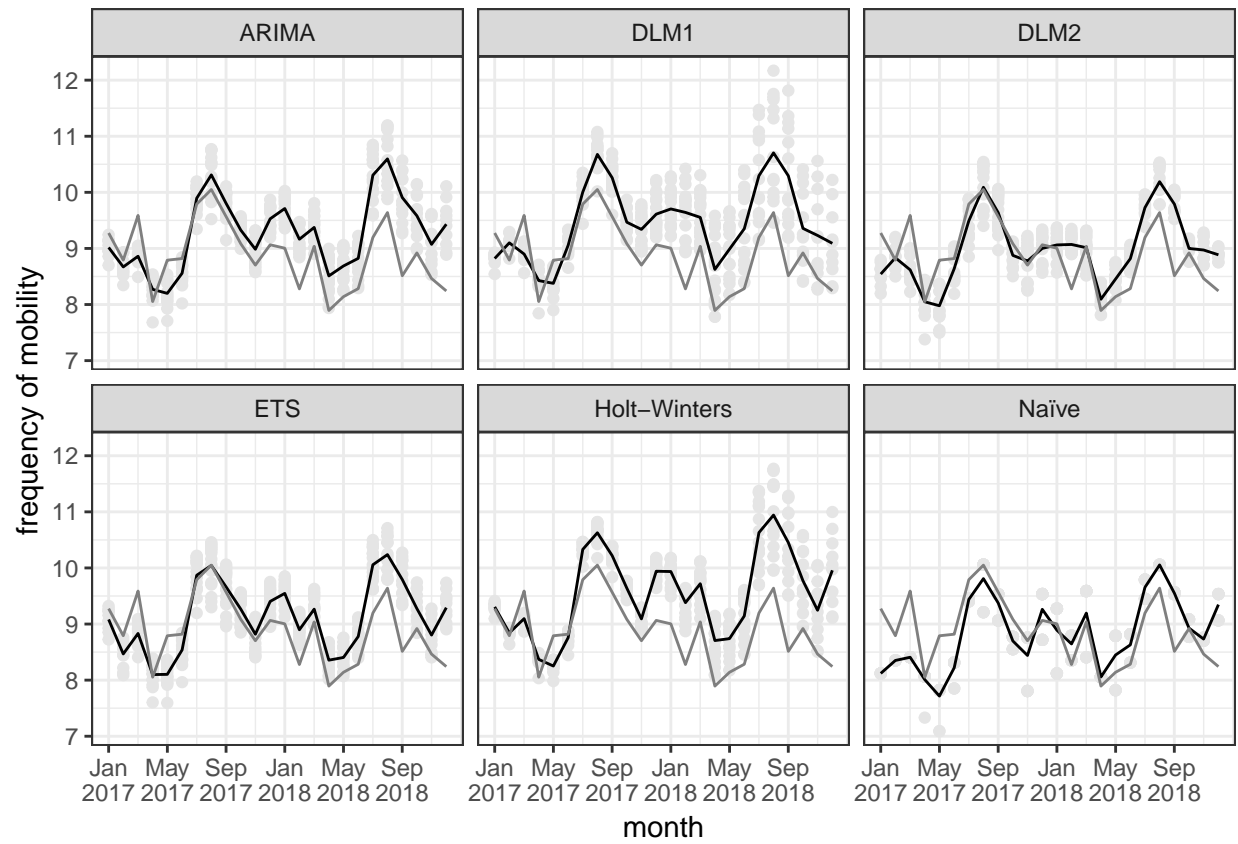


Figure 7: Forecasted (light grey dots), forecast averaged over origins (black line) and observed (dark grey line) frequency of mobility

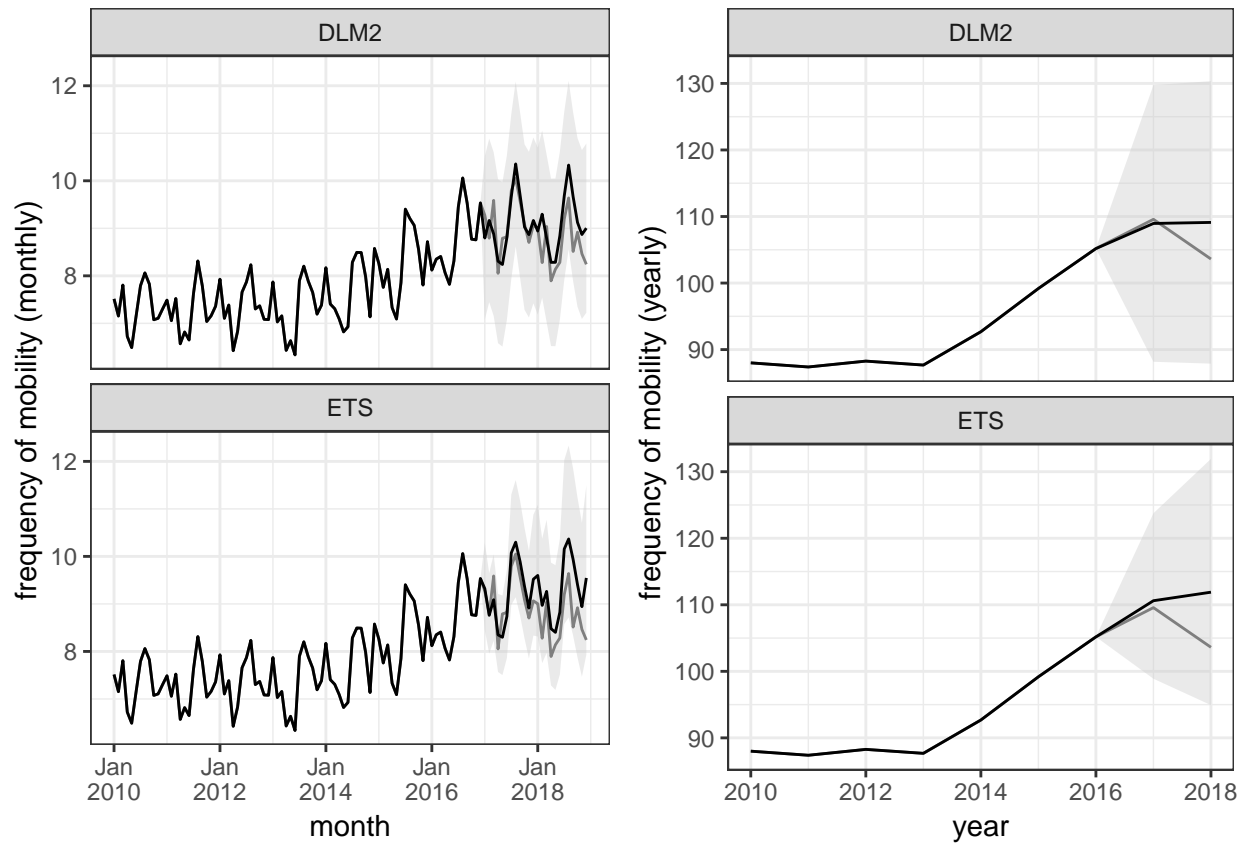


Figure 8: Two-year forecast (black line) of monthly (left) and yearly (right) frequency of mobility and test data (dark grey line) and 95 percent prediction interval

need for this modelling exercise arose as we suspected that the end of our yearly series represented the top of a cycle.

The results presented in the previous section show that the state space model (DLM) quite accurately forecasts both monthly and yearly frequency of mobility until end 2018, based on data up to end 2016. In particular, the model would have given us a very accurate picture of the development of the frequency of mobility until end of 2017. 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.

The most important reason was that the local linear growth component ensured a certain “conservatism” for forecast horizons above 12 months. The high signal-to-noise rate implied that the estimated local growth rate evolved slowly over time. The only other model mimicking this behaviour was the ETS model - itself also a state space model. An easy critique of this result 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. As such, this model also answers the question “come down to what?”.

Although the focus of this paper was univariate forecasts, one of the biggest advantages of the structural model developed in this paper, relative to automatically selected models such as ETS, is the easy with which covariates can be included. Future research should focus on investigating whether model performance can be improved by including covariates for which there are reliable short- to medium-term forecasts - for example unemployment levels. Splitting up the time series into several age groups (dynamic hierarchical models) will probably make such an exercise easier, and is probably a fruitful venue for future research. A more theory-driven model with covariates could also provide extra information about the merits of the simpler univariate approach presented.

7 Appendix A

In addition to the cycle, there are strong seasonal patterns that seem to vary over time. The seasonal subseries plot in Figure 9 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 10 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 ‘scaloped’ 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 (R. J. Hyndman and Athanasopoulos 2018).

8 Appendix B

Bibliography

- Bergmeir, Christoph, and José M Benítez. 2012. “On the Use of Cross-Validation for Time Series Predictor Evaluation.” *Information Sciences* 191. Elsevier: 192–213.
- Bergmeir, Christoph, Rob J. Hyndman, and Bonsoo Koo. 2018. “A Note on the Validity of Cross-Validation

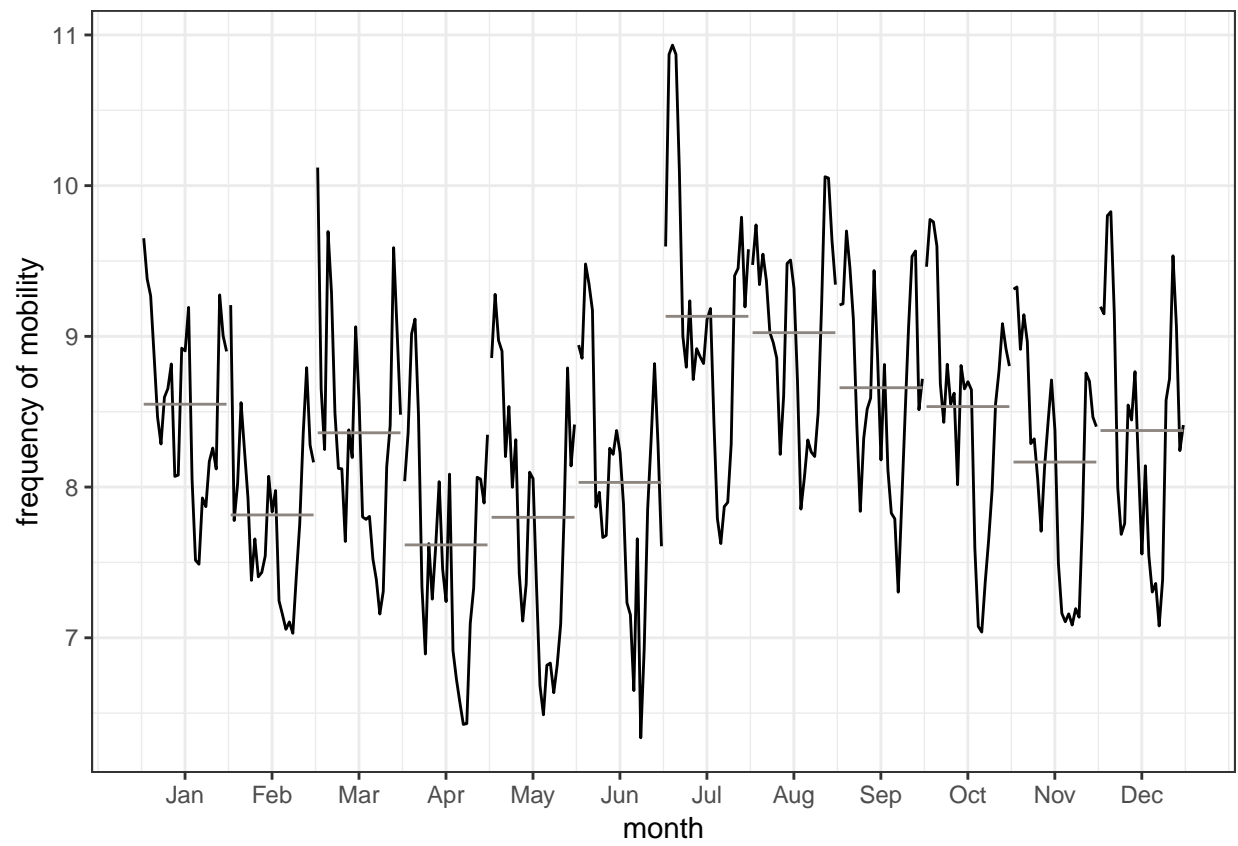


Figure 9: Seasonal subseries plot of the frequency of mobility

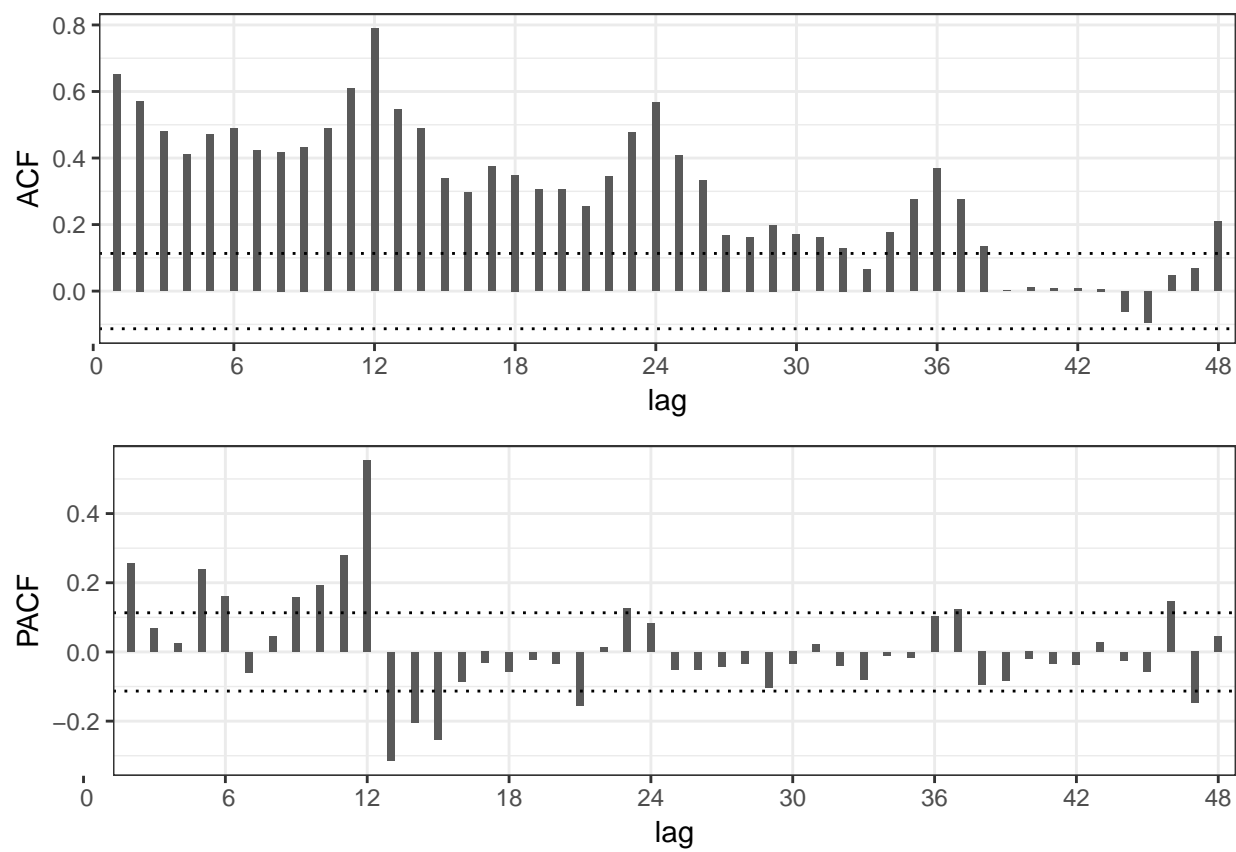


Figure 10: Autocorrelation function of the frequency of mobility

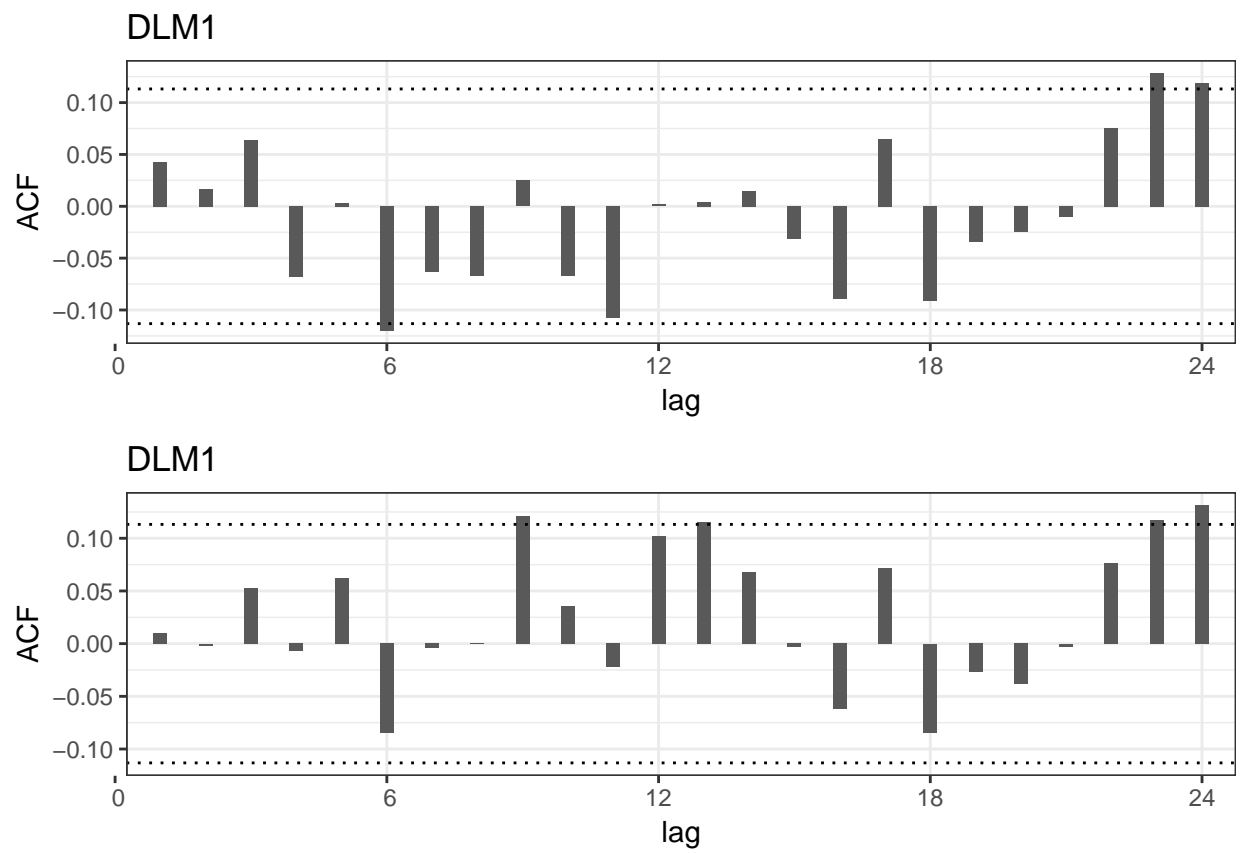


Figure 11: Autocorrelation function of the standardised residuals

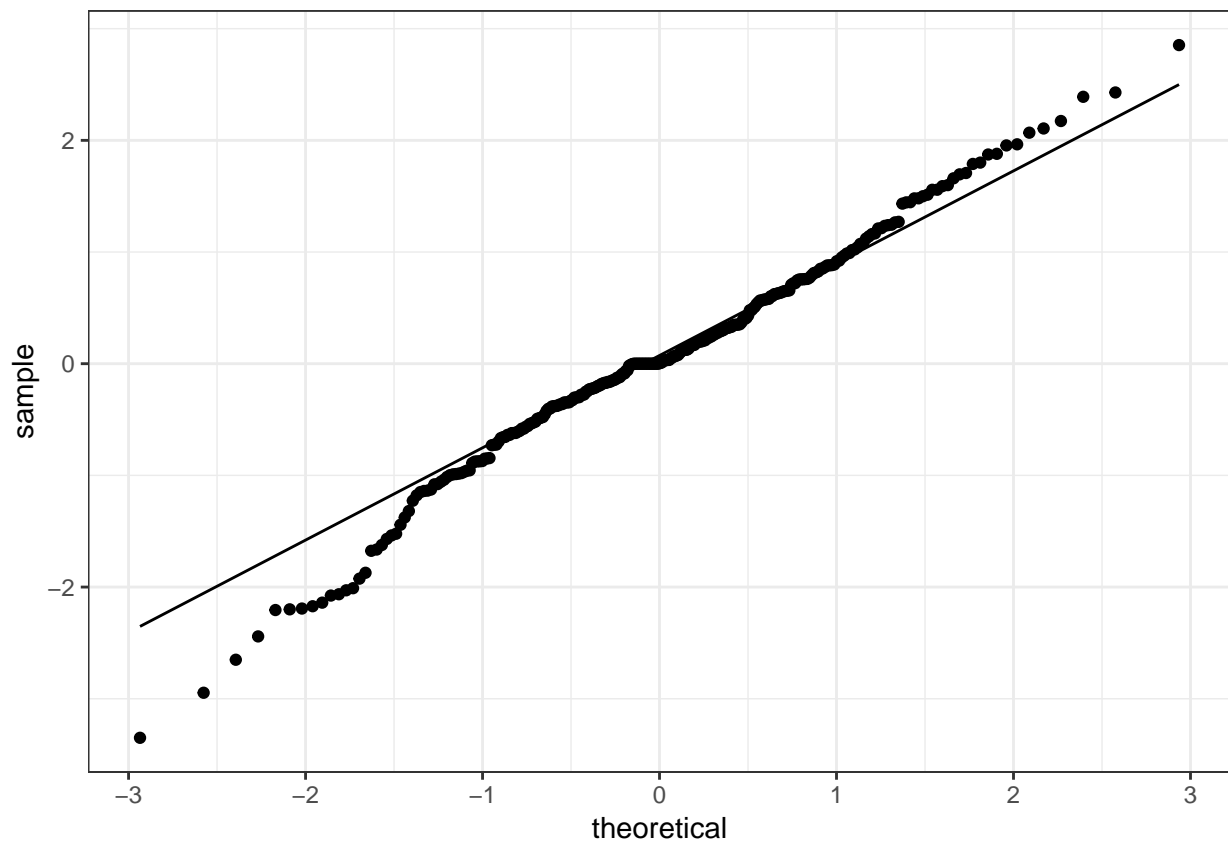


Figure 12: Normal probability plot of standardized one-step-ahead forecast errors of DLM2

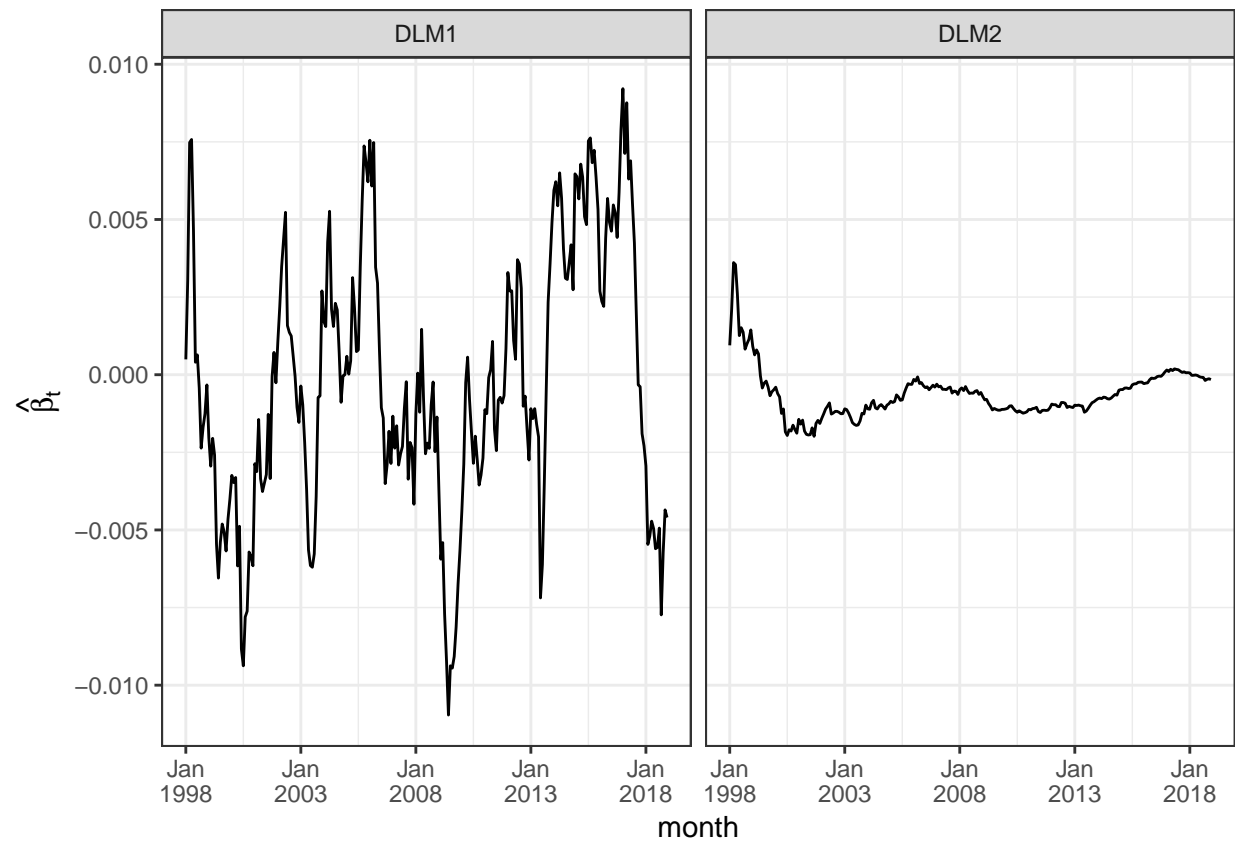


Figure 13: Filtering state estimates of the slope (β_t)

Table 3: Mean and standard deviation of the estimated parameters

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

for Evaluating Autoregressive Time Series Prediction.” *Computational Statistics & Data Analysis* 120: 70–83. doi:<https://doi.org/10.1016/j.csda.2017.11.003>.

Canova, Fabio. 1998. “Detrending and Business Cycle Facts.” *Journal of Monetary Economics* 41 (3). Elsevier: 475–512.

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). MIT Press: 831–43.

Holt, Charles C. 2004. “Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages.” *International Journal of Forecasting* 20 (1). Elsevier: 5–10.

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. <http://www.jstatsoft.org/article/view/v027i03>.

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). Elsevier: 439–54.

Kalman, Rudolf Emil, and others. 1960. “Contributions to the Theory of Optimal Control.” *Bol. Soc. Mat. Mexicana* 5 (2): 102–19.

Kaplan, Greg, and Sam Schulhofer-Wohl. 2017. “Understanding the Long-Run Decline in Interstate Migration.” *International Economic Review* 58 (1). Wiley Online Library: 57–94.

Monahan, John F. 1984. “A Note on Enforcing Stationarity in Autoregressive-Moving Average Models.” *Biometrika* 71 (2). Oxford University Press: 403–4.

Mulder, Clara H. 2018. “Putting Family Centre Stage: Ties to Nonresident Family, Internal Migration, and Immobility.” *Demographic Research* 39. JSTOR: 1151–80.

Petris, Giovanni, Sonia Petrone, and Patrizia Campagnoli. 2009. *Dynamic Linear Models with R*. Springer.

Tashman, Leonard J. 2000. “Out-of-Sample Tests of Forecasting Accuracy: An Analysis and Review.” *International Journal of Forecasting* 16 (4): 437–50. doi:[https://doi.org/10.1016/S0169-2070\(00\)00065-0](https://doi.org/10.1016/S0169-2070(00)00065-0).

te Riele, Saskia, Huisman Corina, Lenny Stoeldraijer, Andries de Jong, Coen van Duin, and Trond Husby. 2019. “PBL/Cbs Regionale Bevolkings- Enhuishoudensprognose 2019–2050: Belangrijkste Uitkomsten.” *Statistische Trends*. CBS.

Winters, Peter R. 1960. “Forecasting Sales by Exponentially Weighted Moving Averages.” *Management*

Science 6 (3). INFORMS: 324–42.

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). Elsevier: 271–81.