



# DEMOGRAPHIC RESEARCH

*A peer-reviewed, open-access journal of population sciences*

---

## ***DEMOGRAPHIC RESEARCH***

**VOLUME 45, ARTICLE 28, PAGES 871–902  
PUBLISHED 29 SEPTEMBER 2021**

<http://www.demographic-research.org/Volumes/Vol45/28/>

DOI:10.4054/DemRes.2021.45.28

*Research Article*

### **Short- to medium-run forecasting of mobility with dynamic linear models**

**Trond Husby**

**Hans Visser**

© 2021 *Trond Husby & Hans Visser.*

*This open-access work is published under the terms of the Creative Commons Attribution 3.0 Germany (CC BY 3.0 DE), which permits use, reproduction, and distribution in any medium, provided the original author(s) and source are given credit.*

*See <https://creativecommons.org/licenses/by/3.0/de/legalcode>*

# Contents

1	Introduction	872
2	Data and modelling	874
2.1	Monthly time series of mobility	874
2.2	Model description	876
3	Results	878
3.1	Parameter estimates	878
3.2	Filtering and smoothing estimates of mobility	879
4	Discussion	882
4.1	Model assessment	882
4.2	Models used for comparison	884
4.3	Empirical results	885
4.4	Forecast values from all origins	887
4.5	Nowcasting and forecasting from 2017 onwards	888
5	Conclusion	891
	References	893
	Appendices	897

## **Short- to medium-run forecasting of mobility with dynamic linear models**

**Trond Husby<sup>1</sup>**

**Hans Visser<sup>2</sup>**

### **Abstract**

#### **BACKGROUND**

Long-term projections of mobility are key inputs to sub-national population projections. These long-term projections are based on extrapolations of long-term trends. In cases of strong, potentially temporal, fluctuations it is informative to analyse the short- to medium-term dynamics of mobility, using data of monthly frequency.

#### **OBJECTIVE**

We develop two univariate models to forecast short- to medium-term mobility in the Netherlands. We apply a recent turning point in the time series of mobility to demonstrate how short- to medium-term forecasts can provide early warning signals about possible changes in the annual trend.

#### **METHODS**

The models we apply are Dynamic Linear Models (DLMs) which belong to the state space family of models. The two models developed in the paper incorporate trend, seasonal and autoregressive components but differ in the representation of the long-term trend. Posterior sampling allows for calculation of consistent prediction intervals for both monthly and annual data.

#### **CONCLUSION**

Forecast accuracy is evaluated using time series cross-validation. Point forecast errors and calibration of prediction intervals are compared to those of several other popular univariate forecasting models. One of our DLM models is more accurate than the models included as comparison.

---

<sup>1</sup> PBL Netherlands Environmental Assessment Agency, the Hague, the Netherlands.  
Email: [trond.husby@pbl.nl](mailto:trond.husby@pbl.nl).

<sup>2</sup> PBL Netherlands Environmental Assessment Agency, the Hague, the Netherlands.

## **CONTRIBUTION**

The paper shows how short- to medium-term forecasts of mobility can be used to inform long-term projections based on annual data. This will be a challenging task for statistical offices generating post-COVID-19 demographic projections.

## **1. Introduction**

Long-term demographic projections based on the cohort-component model require trend extrapolation of growth components (Smith, Tayman, and Swanson 2013). This entails isolating a long-term trend in historical data, using a variety of smoothing- and filtering-methods. It is often useful to complement this exercise with a short- to medium-run forecast. 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). One of the problems with annual time series is that they are released once a year, and the top of a cycle may lie somewhere in the middle of that year. This may be a real challenge facing statistical offices working with post-COVID-19 demographic projections. In this paper we develop methods for short- to medium-term forecasting, evaluating the accuracy of point forecasts and the calibration of intervals, on a time series of monthly mobility in the Netherlands. We use a recent turning point in the time series of mobility to demonstrate how an analysis of the short- to medium-term dynamics can provide early warning signals about possible changes in the annual trend.

As is the case in most developed countries, one of the main drivers of regional-level population change in the Netherlands is mobility. Extrapolations of mobility therefore play a key role in regional population projection (te Riele et al. 2019). Due to the close relationship between mobility decisions and labour- and housing-market conditions, time series of mobility often follows the macroeconomic cycle. However, disentangling the effects of macroeconomic drivers on mobility is not straight forward. 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 economic considerations may, at least in the long run, be just as important in explaining migration. For example, family considerations are important determinants of mobility on an individual level, which means that mobility decisions are affected by changes to family composition and aging (Mulder 2018). In order to forecast such time series, relatively simple univariate methods can be of great value (Zietz and Traian 2014).

Univariate forecasting often involves decomposing a time series into elements such as trend, cycle and seasonality, where these basic time series patterns are used to form a forecast. One model type frequently used for such analyses is the Dynamic Linear Model

(DLM), which belongs to the family 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 filter algorithm. One advantage of these models is that their representation as stochastic systems allows for efficient sampling of the posterior distribution of estimated model parameters. Another advantage is that the models are easily extended: time series components can easily be added to or removed from a model if deemed necessary.

The demographic literature contains a number of applications of state space models based on the Kalman filter. In a similar vein as our paper, de Beer (1988) applies a structural time series model to forecast the growth components of a cohort-component model. Rueda and Rodríguez (2010) introduce multivariate state space models for estimating and forecasting fertility rates. Furthermore, the Kalman filter has been used for other purposes than forecasting. Ordorica-Mellado and García-Guerrero (2016) develop a method, based on the Kalman filter, for dynamic small area estimation of the population. Lee and Anderson (2002) examine the macro-level causes and consequences of population change, setting up a Malthusian system on a state space form where the relevant parameters are estimated with the Kalman filter. Building on the Gaussian state space formulation of the Lee–Carter model, Li et al. (2019) model the evolution of Chinese mortality over time, using the Kalman filter to handle missing data points.

Although the connection is not always acknowledged in the literature, both the Kalman filter itself and the state space formulation clearly draw on Bayesian methods (West and Harrison 2006). As such, this paper relates to a growing literature using Bayesian methods for demographic projections (Bijak and Bryant 2016). See, for example, Bijak (2010) for an overview of Bayesian methods for forecasting international migration. In a series of papers, Raftery and colleagues propose generic models for probabilistic forecasts of the world population (Raftery et al. 2012; Azose and Raftery 2015; Azose, Ševčíková, and Raftery 2016; Azose and Raftery 2019). Wiśniowski et al. (2015) develop dynamic Bayesian models to forecast population, where Lee-Carter models are used to forecast age-patterns of components. In terms of general mobility, one example is Congdon (2000), who uses a Bayesian version of the Gravity Model to forecast patient flows to hospitals.

The theme of our paper – namely forecasting mobility – relates to a broader literature on measuring and comparing internal migration and mobility. One challenge for time series of migration and mobility is that moving is, by definition, a spatial process (Fotheringham and O’Kelly 1989). There are several fundamental issues with measuring migration across time and space (Bell et al. 2002). One problem is related to the granularity of the spatial units: changing granularity changes the observed intensity of mobility – the so-called modifiable areal unit problem (Fotheringham and Wong 1991). A related problem with long time series is changes to administrative borders, where adjustments have to be made to achieve consistency (Husby et al. 2014). Bijak et al. (2019) carry

out an empirical comparison of methods for forecasting (international) migration. They recommend a three-step process for migration modelling: (1) understand the features of the particular migration flow; (2) assess the available data; and (3) select a modelling approach appropriate for the type of migration and the available data.

In this paper we develop two DLMS for a short- to medium-term forecast of mobility, where short- to medium-term is understood as a time horizon up to two years. In particular, we are interested in whether these models provide reasonable short- to medium-term forecasts around a turning point in a time series. We also illustrate how both point forecasts and intervals of monthly series can be aggregated to annual frequencies. As a robustness check we compare the forecasting performance of our model with that of six other popular models for univariate time-series forecasting. These models are all estimated using automatic routines available in *R*. 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 structure of the paper is as follows: Section 2 discusses the data and presents the model; Section 3 shows the estimated parameters and results from basic inference with the Kalman filter; Section 4 presents results from an out-of-sample evaluation and illustrates how forecasts of monthly mobility can inform forecasts on an annual scale; and Section 5 summarises the findings and concludes the paper.

## 2. Data and modelling

### 2.1 Monthly time series of mobility

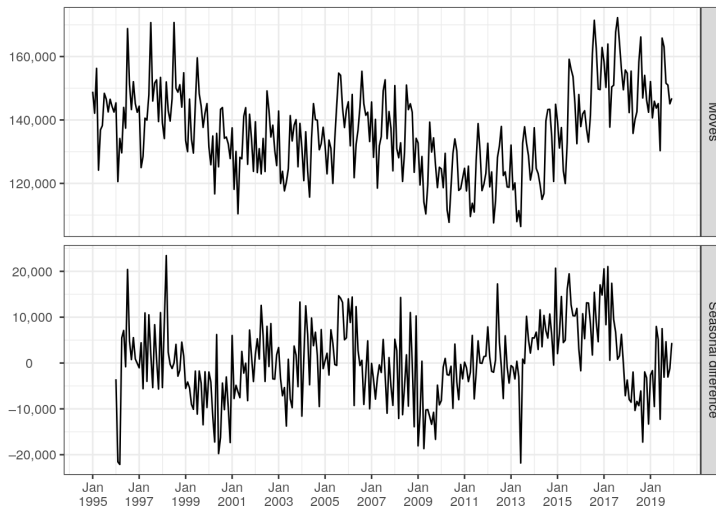
The national mobility rate is a key variable to the cohort-component model used for the regional population projections of Netherlands Environmental Assessment Agency (PBL) and Statistics Netherlands (CBS), namely the Projecting population Events At Regional Level (PEARL, see de Jong et al. 2005). In terms of internal migration, PEARL is a bottom-up model with an explicit representation of origin-destination flows between municipalities and within-municipality moves. Trend extrapolation of the national mobility rate is used to ensure consistency between the municipality- and national-level of mobility. As such, the national mobility rate has an important controlling mechanism in the overall model. Since the national population hardly changes in the short- to medium-run, the short- to medium-term dynamics in the mobility rate stem almost entirely from fluctuations in mobility counts. For this reason, we focus on time series of mobility counts rather than rates.

The time series applied in this paper is the national level of mobility in the Netherlands, which is an aggregate of recorded address changes in the population register of the CBS. The data can be classified as 'events' (Bell et al. 2002), including both inter- and

intramunicipal moves. Since we use national aggregates of mobility counts, the data are not prone to the modifiable areal unit problem. As stressed by Bijak et al. (2019), the variability of migration data can be a major source of uncertainty in migration forecasts. In the context of mobility there could be biases for certain groups, such as students who stay registered at their parents' home. However, since the data covers the entire population, this is unlikely to be a serious concern. Furthermore, the data used for the time series in the paper has been subject to ex-post corrections by the CBS.

Using open data from the CBS, we obtain a monthly time series of moves from January 1995 to December 2019, leading to a total of 300 observations. More formally, we define the mobility in time period (year and month)  $t$ ,  $m_t$ , as the sum of inter and intramunicipality moves between the first and the last day of the month in  $t$ . Figure 1 shows the time series from January 1995 to December 2019 as well as the seasonal differences, defined as  $m_t - m_{t-12}$ .

**Figure 1: Mobility (upper panel) and seasonal differences (lower panel)**



The upper panel shows that there are between 110,000 and 170,000 moves per month, which corresponds to between 7 and 11 moves per 1,000 inhabitants per month. The plot suggests that the mobility is cyclical with an uneven period length. It is clearly seen from the plot that the time series is nonstationary, which is also confirmed by a Dickey Fuller test. The financial crisis, which hit the Dutch housing market hard, can be

seen in the dip from 2008 onward, with the recovery from the crisis set in during the year 2014.

From the bottom panel, which shows the seasonal differences, we see that the year-to-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 reveals that the year-to-year changes become negative from mid 2017. This turning point reflects an end of the increasing trend from 2014.

Summing up, the decomposition of the time series suggests that there is no linear trend, rather a cycle with peaks roughly every ten years. Furthermore, there are strong seasonal effects with variations between the years, indicating a noisy seasonality component. In addition, the seasonality and the trend-cycle lead to autocorrelation.

## 2.2 Model description

State space models are a representation of a dynamic system, where observed values are a linear function of an unobserved process (the state) and noise. The state space formulation is a necessary condition to carry out inferences by the Kalman filter algorithm (Kalman 1960). The filter is computational recursive, allowing for estimates of the state in a way that minimizes the mean of the squared one-step-ahead prediction errors or innovation. The filter supports estimates of past, present, and even future states. As mentioned above, DLMS form a special case of state space models where errors are normally and independently distributed.

Let the time series  $y_t$  denote the logarithm of mobility. In state space formulation,  $y_t$  is a linear function of a latent underlying vector of states  $\theta_t$ . Furthermore,  $y_t$  are conditionally independent given the state  $\theta_t$ , 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:

$$y_t = F\theta_t + v_t \qquad v_t \sim N(0, V_t) \qquad (1)$$

$$\theta_t = G\theta_{t-1} + w_t \qquad w_t \sim N(0, W_t) \qquad (2)$$

$$\theta_0 \sim N(m_0, C_0). \qquad (3)$$

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 matrix  $V_t$  and system variances in the matrix  $W_t$ .  $F$  and  $G$  are known system matrices. We use noninformative priors for  $\theta_0$ , with mean  $m_0$  and variance matrix  $C_0$ .

Elements of the state vector  $\theta_t$  should be chosen so that they reflect the characteris-



tics of the time series. As noted in the previous section, 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  $\mu_t$  which varies over time with slope  $\beta_t$ . The dynamics of the slope itself is modelled as a random walk. We can write the local linear trend model as follows:

$$y_t = \mu_t + v_t \quad v_t \sim N(0, \sigma_v^2) \quad (4)$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + w_{t,1} \quad w_{t,1} \sim N(0, \sigma_\mu^2) \quad (5)$$

$$\beta_t = \beta_{t-1} + w_{t,2} \quad w_{t,2} \sim N(0, \sigma_\beta^2). \quad (6)$$

Seasonality is dealt with by the inclusion of a trigonometric seasonal components 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 smooth seasonal pattern. In 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 over time. For the  $j$ th harmonic we can write the evolution of the seasonal effects as:

$$S_{j,t+1} = S_{j,t} \cos \omega_j + S_{j,t}^* \sin \omega_j \quad (7)$$

$$S_{j,t+1}^* = -S_{j,t} \sin \omega_j + S_{j,t}^* \cos \omega_j. \quad (8)$$

with the Fourier frequencies  $\omega_j = \frac{2\pi t j}{s}$ . We include two harmonics.

The two components discussed thus far, trend and seasonal, remove much of the autocorrelation. However, a Ljung–Box test confirmed that the residuals of such a model are still not white noise. One way of dealing with residual autocorrelation is to include autoregressive elements (AR) into the state space model.

In order to determine the order of the AR elements, we proceed in a similar way as that of a seasonal ARIMA model, namely by examining the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the time series data. This procedure is described in Box and Jenkins (1976), who provide both a theoretical framework and practical rules for determining appropriate number of lags. The ACF and PACF, as well as an examination of the resulting model residuals, suggest an AR (12) where only four of the estimated parameters take non-zero values; namely the parameters of lags 1, 2, 7, and 12. The full procedure is described in detail in Appendix A. Figure B-1 in Appendix B shows the resulting ACF of the standardised residuals, suggesting that the residuals in the first year are essentially white noise. Note that the standard routine for creating autoregressive terms 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.

### 3. Results

Before we apply the Kalman filter for estimation and forecasting, we estimate the unknown variances and the parameters in the autoregressive component using maximum likelihood (see Petris, Petrone, and Campagnoli 2009: Ch. 4). With the estimated parameters we can estimate the remaining parts of the model using the Kalman filter. One important inferential task is to estimate  $\theta_t$  with data up to  $t$ :  $y_1, y_2, \dots, y_t$ . This is referred to as filtering. A second task is smoothing, which entails estimating  $\theta_t$  with all available data:  $y_1, y_2, \dots, y_t, \dots, y_s$ . This section presents the parameter estimates and the resulting filtering and smoothing estimates of mobility.

#### 3.1 Parameter estimates

As shown in the previous section, there are 10 potentially non-zero parameters that need to be estimated before running the Kalman algorithm:  $\sigma_v^2$ ,  $\sigma_\mu^2$ ,  $\sigma_\beta^2$ ,  $\sigma_{S_1}^2$ ,  $\sigma_{S_2}^2$ , and  $\sigma_u^2$ , as well as the autoregressive parameters  $\phi_1$ ,  $\phi_2$ ,  $\phi_7$ , and  $\phi_{12}$ . Note that  $\phi_1$ ,  $\phi_2$ ,  $\phi_7$ , and  $\phi_{12}$  are estimated subject to stationarity restrictions (Monahan 1984). Initial analyses revealed that the value of  $\sigma_{S_1}^2$  hardly had any discernible impacts on model results, therefore this parameter is set to zero.

The first two parameters in  $W$ ,  $\sigma_\mu^2$  and  $\sigma_\beta^2$ , represent the variances for the trend-cycle component; one for the level and one for the slope. Initial runs revealed that two special cases of the model fit the data better than that with both variances larger than zero. The first special case, sometimes referred to as the integrated random walk or the smooth trend model (Young et al. 1991), is obtained by setting  $\sigma_\mu^2 = 0$ . In this specification, all noise patterns are shifted from the level component to the slope, meaning that the trend-cycle component of the state vector becomes smooth (hence the name). In the remaining parts 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. Here the estimated  $\beta_t$  reflect the long-run slope of the time series rather than the local slope, essentially reducing to what is sometimes called the random walk with drift (Koopman and Ooms 2011). Due to the inclusion of the autoregressive elements, the short-run forecasts can diverge from long-term growth. More specifically, a one-step-ahead prediction  $\hat{y}_{t+1}$  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.

Estimated parameter values are shown in Table 1, along with standard errors calcu-

lated using the Delta method. The observation variance  $\sigma_v^2$  is, for all practical purposes 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. The variance of the seasonal component,  $\sigma_{S_2}^2$ , is larger than zero for both models, allowing seasonality to change over time.

**Table 1: Estimated parameters and standard errors**

Parameter	DLM1 Estimate	SE	DIM2 Estimate	SE
$\sigma_v^2$	0.00000002	0.000	0.00000002	0.000
$\sigma_{\mu}^2$			0.0001	0.00003
$\sigma_{\beta}^2$	0.000002	0.000001		
$\sigma_{S_2}^2$	0.000008	0.000004	0.000008	0.000004
$\sigma_u^2$	0.0013	0.0001	0.0010	0.0001
$\phi_1$	-0.637	0.2356	-0.6904	0.2311
$\phi_2$	0.0308	0.1221	-0.0537	0.1274
$\phi_7$	-0.4415	22.0938	-0.4879	0.0953
$\phi_{12}$	0.4486	0.0623	0.4349	0.0616

### 3.2 Filtering and smoothing estimates of mobility

With the Kalman filter we obtain the filtered distribution of states 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  $\theta_{t+1}$  and then predicting  $\hat{y}_{t+1}$ . Similarly, an  $n$ -step ahead forecast  $y_{t+n}$  is based on calculating the  $n$ -step ahead state vector  $\theta_{t+n}$ . We can write the forecast function as  $f_t = E(y_{t+n}|y_1, y_2, \dots, y_t)$ . Besides the filtering distribution, we also estimate a smoothing distribution, representing the states conditional on the observed values of the entire time series  $\theta_t|y_1, y_2, \dots, y_s$ , where  $s \geq t$ .

The one-step ahead predictions, or filtering estimates, from both of the models are displayed in Figure 2 together with the Mean Absolute Percentage Error (MAPE). Due to the recursive nature of the Kalman algorithm, there is a large discrepancy between the filtered values and the observed values in the first periods. We therefore discard the three first years when calculating the MAPE. The MAPE suggests that DLM2 fits the data somewhat better than DLM1, though differences are not dramatic. A residual check showed that the normality assumption of DLM2 is justified (see Appendix B).

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

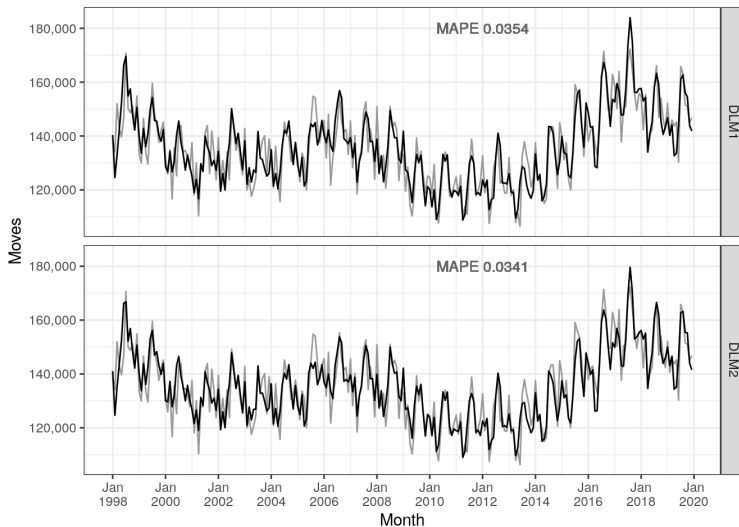
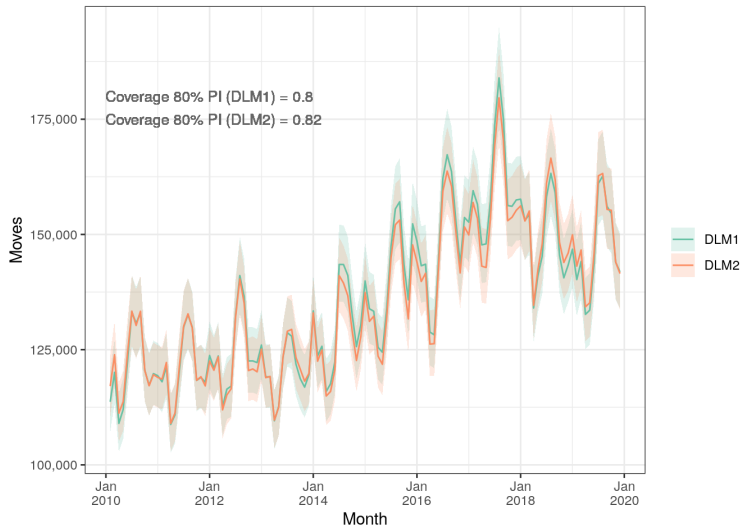


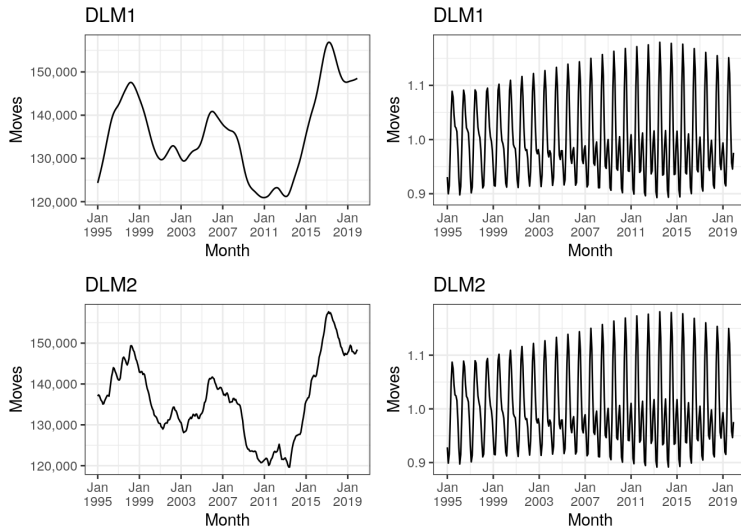
Figure 3 shows the one-step ahead Kalman filter forecasts along with 80% prediction intervals of DLM1 and DLM2 for the months between January 2010 and December 2018. The figure also reports on the interval coverage probability, calculated as the share of predicted values falling within the interval. 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 mobility than DLM2 from 2014 to about 2017.

The smoothing estimates for the trend-cycle component and seasonal effects are shown in Figure 4. From the left panels we can pinpoint the recent peak in the trend-cycle to March 2017. There is 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 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 the development of the seasonal factors from the raw data well.

**Figure 3: Filtering estimate and 80% prediction intervals**



**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. We evaluate out-of-sample point forecasts and interval coverage using standard metrics, comparing the accuracy to that of several other univariate forecasting models. In particular we focus on a turning point in the time series, evaluating how forecasting errors vary as a function of the length of the forecast horizon.

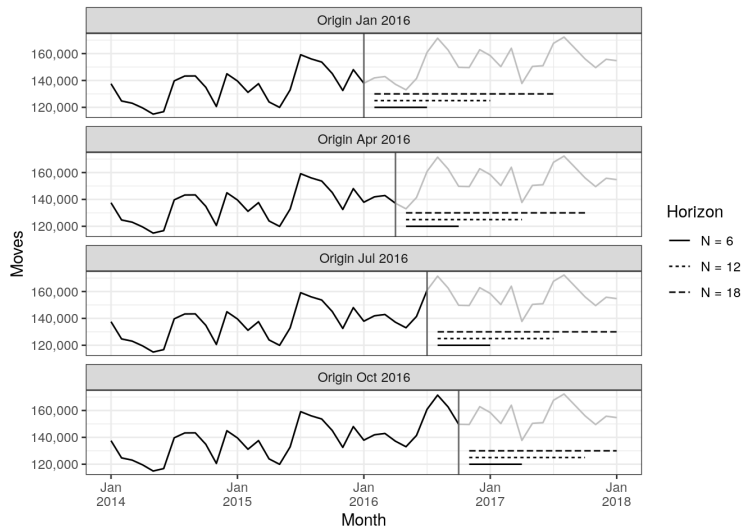
### 4.1 Model assessment

Model evaluations on cross sectional data are often 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 for varying training and test data sets. This procedure is called k-fold cross validation (Hastie, Tibshirani, and Friedman 2009). In the time series literature, the test data generally do 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 Markov assumption embedded in the Kalman filtering algorithm, we expect that standard cross-validation is not appropriate for the DLMS in this paper.

We follow the conventional approach for evaluating forecast errors, namely evaluation over a rolling origin (Tashman 2000), where the training data are successively extended in  $k$  iterations. The last point of the training data in each iteration is referred to as the origin  $T$ , meaning that 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. The model is estimated on the training data (black solid line) and forecast performance is evaluated by averaging the forecast errors over different horizons (dotted horizontal lines).

We evaluate forecasting performance over three time horizons:  $N = 6$ ,  $N = 12$  and  $N = 18$ . This means that there are three versions of each error measure: one for a half year forecast, one for a year, and one for one and a half years. We are interested in assessing the performance around the change point in March 2017 (see Figure 4). In order to evaluate the performance of the models on both sides of the change point, we include origins within 12 months before and after March 2017. Including the change point itself, we arrive at a total of 25 forecast origins:  $T =$  March 2016, February 2016,  $\dots$ , and June 2017.

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



Point predictions are evaluated using MAPE and the Mean Absolute Scaled Errors (MASE), both averaged over forecast origins. MAPE is included as it is widely used and easily interpreted. However, MAPE has a number of potential drawbacks, including putting a heavier penalty to negative than to positive forecast errors (Hyndman and Koehler 2006). MASE overcomes some of the issues of MAPE and has a simple interpretation. The seasonal MASE is defined as the mean absolute forecast error divided by the mean absolute (within-sample) error of the one-step-ahead seasonal naïve forecast. Consequently, a MASE below 1 indicates smaller forecast errors than a one-step-ahead naïve forecast.

The calibration of both 80% and 95% prediction intervals is assessed with interval coverage probabilities – a standard measure of model error within stochastic population forecasting (Bijak et al. 2019). The coverage probability of the intervals is calculated as the proportion of the observations between March 2016 and June 2017 that fell within their prediction intervals (PI). In well-calibrated models it is expected that the empirical frequencies are close to the nominal coverage probabilities.

## 4.2 Models used for comparison

The forecasting accuracy of the DLMS is compared to the accuracy 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 short-term forecasts of 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 change point and increasingly worse around and after it.

Furthermore, we include general exponential smoothing models on state space form, referred to as Error, Trend, Seasonal (ETS). The state space formulation allows all combinations of trend-, seasonal and error components to be modelled (Hyndman et al. 2002). The automatic model selection is carried out using the `ets()` function with default values for all arguments, and model selection is based on the corrected Akaike’s Information Criterion (AIC).

We also include a basic variant of the Structural Time Series model described in Harvey (1990). 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 monthly dummies to control for seasonality.

Another widely used model, the so-called Trigonometric seasonality, Box-Cox transformation, ARMA errors, Trend, and Seasonal (TBATS) components model, extends the ETS model by allowing for multiple seasonalities (De Livera, Hyndman, and Snyder 2011). This model is also implemented as a state space model, and its components are automatically selected using a similar routine as that of the ETS model. One reason for including this model is that the trigonometric seasonality, also used in the DLMS, filters out some of the noise in the time series.

Finally, we include a seasonal ARIMA model in 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 a model of the form ARIMA (2,1,5) (2,1,1). This is a model with double differences, with yearly and monthly MA (2), and yearly AR (5) and monthly AR (2). For each origin, we re-estimate the model parameters on the respective training data. We also tried the automatic selection for each origin with virtually identical results.



### 4.3 Empirical results

Figure 6 shows the overall forecast accuracy across origins. From the figure we see that the point forecasts of DLM2 outperform DLM1 for all forecast horizons, and especially for 18 months forecast. For the forecast horizon of 6 months, the MASE below 1 suggests that DLM2 outperforms the one-step-ahead Naïve model. Compared to all other models, the point forecasts of DLM2 are very accurate, both in terms of MAPE and MASE. The relative performance of the models is practically the same across the error measures, suggesting that MAPE is a valid accuracy measure in this context.

Not surprisingly, the forecast accuracy of DLM2 deteriorates as  $N$  increases but, compared to the other models, the decrease in accuracy is limited: the mean MAPE for  $N = 6$  is around 3.5%, compared to around 5% for  $N = 18$ . The limited decrease in forecast accuracy is not shared with DLM1, the ARIMA, and Holt–Winters models, which all see substantial increases in forecasting errors as the forecast horizon increases. The only model that shares this feature with DLM2 is the ETS model, which has slightly larger forecast errors for all horizons. Both MAPE and MASE indicate that the DLM2 and ETS models are fairly competitive, with a slight advantage to DLM2, across all forecast horizons. The figure also suggests that DLM1 is among the models with the worst forecast accuracy (with errors increasing with the forecast horizon).

In terms of coverage probability, we see that the empirical forecast distribution of DLM2 matches the theoretical 80% prediction interval closely. The coverage probability of the 95% interval lies in the range between 0.91 and 0.97, depending on the forecast horizon. The empirical 80% intervals of the ETS match the theoretical counterpart fairly well for the shorter forecast horizons. However, its 95% intervals are too conservative for the 18 month forecast horizon, meaning that more than 95% of the observations fall within the theoretical interval. Note that coverage probability of the Naïve and the Holt–Winters models are excluded due to numerical problems with calculating prediction intervals.

**Figure 6: MAPE and MASE of forecasts, coverage probability of forecast intervals**

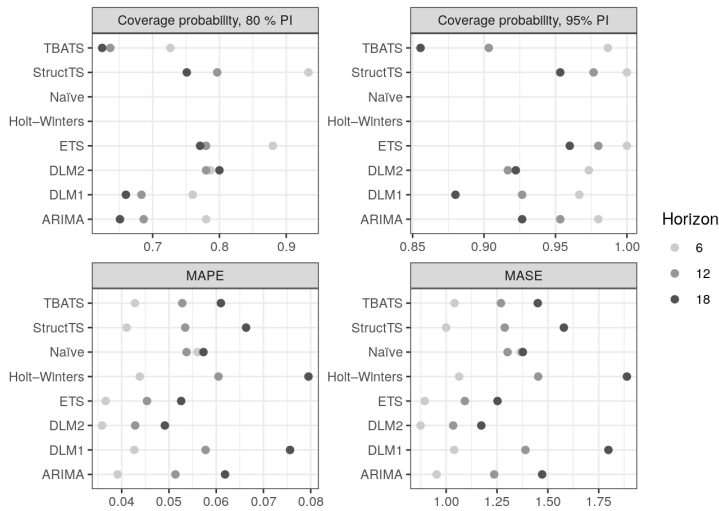
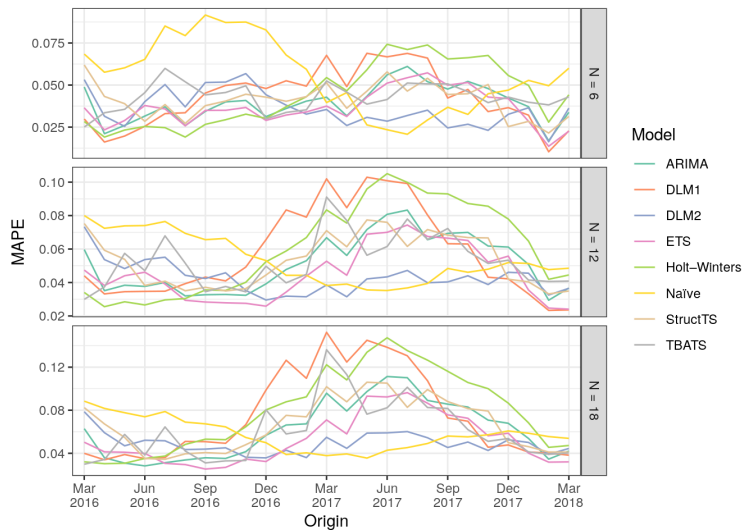


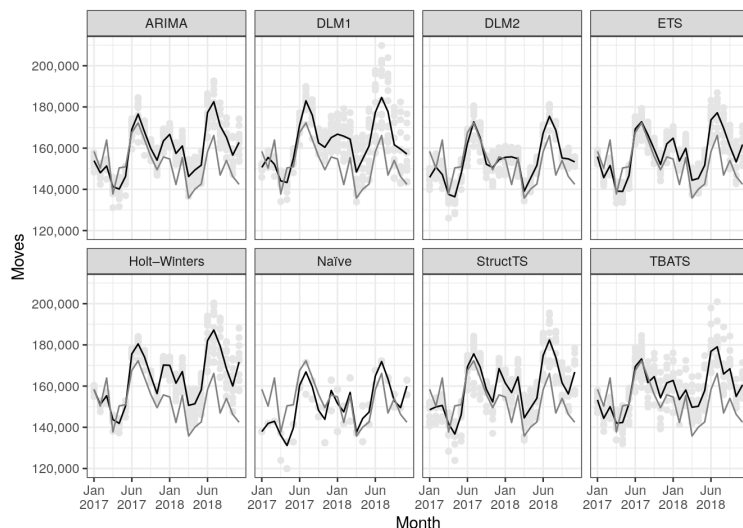
Figure 7 delves further into the variation in forecast performance across origins, showing the MAPE per origin and model for the three forecast horizons. In the interval between December 2016 and December 2017, DLM2 is the best performing model for all forecast horizon, with the exception of origins right after March 2017 where the Naive 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 Figure 6 showed, the higher forecast accuracy of the ETS model, relative to the DLMS, occurs primarily for a short forecast horizon. This is caused by the relatively high MAPE for the DLMS prior to December 2016.

**Figure 7: MAPE by origin and forecast horizon**

#### 4.4 Forecast values from all origins

In addition to general evaluations of forecast performance, we are also interested in the value of some specific forecasts – namely those around the change point. Figure 8 compares the forecasts of all models in the period January 2017 to December 2018. The grey dots represent forecasts from different origins, the black line represents the average value of the forecasts across all origins and the dark grey 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 mobility already from June 2017, meaning that the forecast mobility exceeds the data for all origins. This occurs for the ARIMA model from October 2017 onward. In the case of DLM2, the last forecast point with a value lower than that of the data occurs in May 2018.

**Figure 8:** Forecast of mobility (light grey dots), forecasts averaged over origins (black line), and observed data (dark grey line)



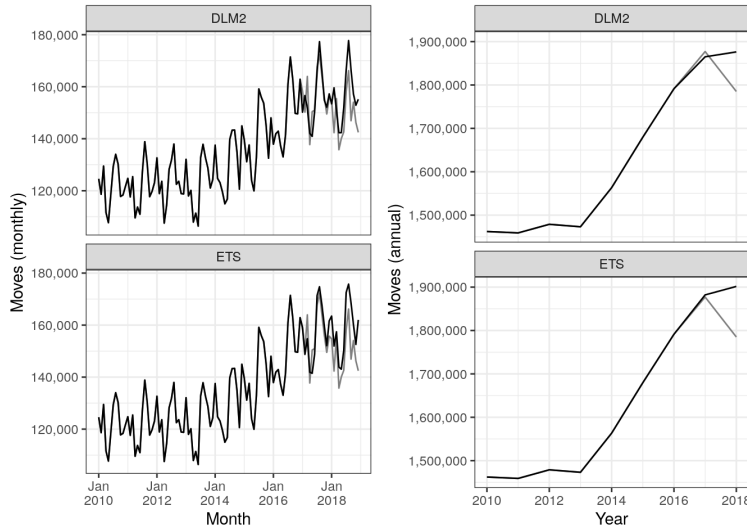
#### 4.5 Nowcasting and forecasting from 2017 onwards

For practical applications of our model it is relevant to know what a model forecast can tell us about annual mobility from March 2017 onward, both in terms of point forecasts and prediction intervals. By applying the rolling origin technique and aggregating the monthly forecasts, we can essentially simulate the annual mobility in 2017 – a technique known as ‘nowcasting’. We can then examine how the forecast annual mobility change as new information becomes available. One potential challenge of this exercise is that the prediction intervals presented in the previous subsections are based on normality assumptions of the residuals on a monthly frequency. The prediction intervals of the monthly series can not be directly translated to an annual frequency. This challenge can be overcome with state space models: by sampling from the posterior distribution of the state vector and aggregating simulated monthly values into annual frequency, we can calculate a Monte Carlo approximation of the intervals of annual mobility.

As suggested by the results of the previous subsections the best performing model, next to the DLM2 model, was the ETS model, which also allows for simulation of future sample paths. The automatic selection procedure results in an ETS (M, Ad, A): a model with multiplicative errors, damped additive trend and additive seasonal effects.

To begin, we simply aggregate the mean forecasts. The left panels in Figure 9 show two-year forecasts of the monthly mobility, while the right panels show mobility on annual frequencies.

**Figure 9: Two-year forecast (black line) of monthly (left) and annual (right) mobility and test data (dark grey line) and 95% prediction interval**

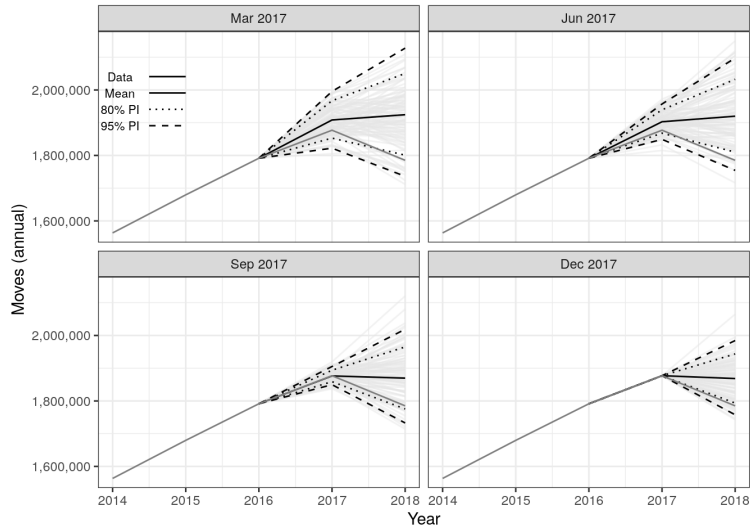


At first glance, the point forecasts of the monthly frequency (left part of the figure) are quite similar. Both models forecast mobility in 2017 accurately and they both overestimate mobility in 2018. As the right left panels of Figure 9 show, the annual mobility declines between 2017 and 2018, and none of the models were able to capture this sudden movement. However, we see ETS is much more “positive” about (i.e., overestimates more substantially) the annual mobility 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 suggests that the models produce quite similar short-run forecasts, but they differ as the horizon increases.

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

mobility in 2017. Figure 10 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) for DLM2.

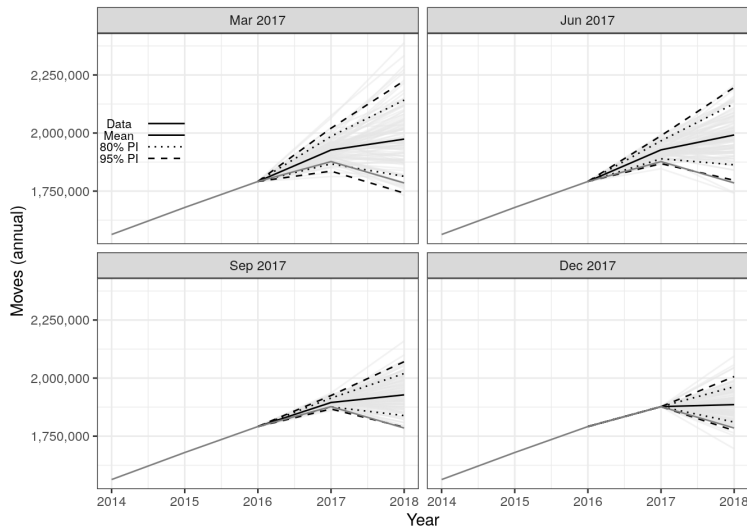
**Figure 10: Monte Carlo forecasts of annual mobility with DLM2 for four different origins**



Since we have identified March 2017 as the change point of the cycle, it is logical that this origin also results in the worst mean forecast (the black line in Figure 10). Moving forward to the origin of June 2017, we can see that the mean forecasts of 2017 and 2018 are closer to the observed data. From September onward we see that the model clearly has picked up the development from after the change point, with the annual mobility declining. The 95% prediction intervals include the observations for all origins, while the 80% prediction intervals include the observed value of 2017 for all origins, and also that of 2018 for the origin of September 2018.

Figure 11 shows the results of the same exercise with the ETS model.

**Figure 11: Monte Carlo forecasts of annual mobility with ETS model for four different origins**



As Figure 9 already suggested, the longer-term forecasts of the ETS diverge from those of the DLM2: the forecasts from the two earliest origins show a stronger increase in mobility between 2017 and 2018. Consequently, the observed mobility in 2018 falls outside the 95% intervals and in some cases even outside the 80% intervals.

## 5. Conclusion

In this paper we have developed two Dynamic Linear Models for short- to medium-run univariate forecasts of monthly mobility in the Netherlands. We have shown how the state space type of models, when interpreted in terms of Bayesian inference, allows one to aggregate monthly point forecasts and intervals to annual frequencies. We paid particular attention to forecast performance around a change point in the time series. In that sense we believe the method presented in this paper can be informative to statistical offices involved in post-COVID-19 population projections. Although our paper deals with a component specific to subnational projections, the method presented here could equally well be used for short- to-medium-run forecasts of components of national projections.

We compared the prediction accuracy of the DLMs to that of a number of other models, estimated using automatic model selection routines. We showed how forecasts

of monthly time of series can be used to generate point forecasts and prediction intervals on an annual frequency, before the annual observations are available. Our paper therefore provides an update to the analysis in de Beer (1988). However, our analysis considers a wider class of models and provides more information about prediction intervals.

The results presented in the previous section show that one of the DLMS quite accurately forecasts both monthly and annual mobility until the end of 2017, based on data up to the end of 2016. More generally, DLM2 performs comparable or better than a number of popular univariate forecasting in a time interval between the end of 2016 to the end of 2017. One remarkable aspect of the forecasts of DLM2 is the relatively limited deterioration in forecast accuracy as the forecast horizon increased. Comparing the two DLMS, we saw that the specification of the local linear growth component played a crucial role. The limited dynamics in the slope term of DLM2 ensures a certain conservatism for forecast horizons above 12 months, thereby reducing the forecast errors. The only other model mimicking this behaviour is the ETS model, which is itself also formulated as a state space model.

Comparing the dynamic linear models of this paper to the Bayesian models described in Bijak and Bryant (2016), one main difference is the use of (unique) model parameters estimated with maximum likelihood. A full Bayesian approach would take into account the parameter uncertainty in predictions and other inference. Whether this matters for the model outcomes is an empirical question. However, for a practitioner this entails some differences in the setup of the model. Although our model results were fairly insensitive to the priors of the state vector, they were quite sensitive to the starting values of the maximum likelihood estimation. In a full Bayesian approach, the task of finding good starting values is replaced by the need to define good priors for the same hyperparameters, which is itself challenging for more complex model types.

Although the time series data used in this paper represent an aggregated measure of mobility, it is well known that migration follows highly age-specific patterns (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 an interesting option for future research.



## References

- Azose, J.J. and Raftery, A.E. (2015). Bayesian probabilistic projection of international migration. *Demography* 52(5): 1627–1650. doi:10.1007/s13524-015-0415-0.
- Azose, J.J. and Raftery, A.E. (2019). Estimation of emigration, return migration, and transit migration between all pairs of countries. *Proceedings of the National Academy of Sciences* 116(1): 116–122. doi:10.1073/pnas.1722334116.
- Azose, J.J., Ševčíková, H., and Raftery, A.E. (2016). Probabilistic population projections with migration uncertainty. *Proceedings of the National Academy of Sciences* 113(23): 6460–6465. doi:10.1073/pnas.1606119113.
- Bell, M., Blake, M., Boyle, P., Duke-Williams, O., Rees, P., Stillwell, J., and Hugo, G. (2002). Cross-national comparison of internal migration: Issues and measures. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 165(3): 435–464. doi:10.1111/1467-985X.t01-1-00247.
- Bergmeir, C. and Benítez, J.M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences* 191: 192–213.
- Bergmeir, C., Hyndman, R.J., and Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics and Data Analysis* 120: 70–83. doi:10.1016/j.csda.2017.11.003.
- Bijak, J. (2010). *Forecasting international migration in Europe: A Bayesian view, vol. 24*. Dordrecht: Springer Science and Business Media.
- Bijak, J. and Bryant, J. (2016). Bayesian demography 250 years after Bayes. *Population Studies* 70(1): 1–19. doi:10.1080/00324728.2015.1122826.
- Bijak, J., Disney, G., Findlay, A.M., Forster, J.J., Smith, P.W., and Wiśniowski, A. (2019). Assessing time series models for forecasting international migration: Lessons from the United Kingdom. *Journal of Forecasting* 38(5): 470–487. doi:10.1002/for.2576.
- Box, G. and Jenkins, G. (1976). *Time series analysis: Forecasting and control*. San Francisco: Holden–Day.
- Canova, F. (1998). Detrending and business cycle facts. *Journal of Monetary Economics* 41(3): 475–512. doi:10.1016/S0304-3932(98)00006-3.
- Congdon, P. (2000). A Bayesian approach to prediction using the gravity model, with an application to patient flow modeling. *Geographical Analysis* 32(3): 205–224. doi:10.1111/j.1538-4632.2000.tb00425.x.
- de Beer, J. (1988). Predictability of demographic variables in the short run. *European Journal of Population/Revue européenne de Démographie* 4(4): 283–296.

[doi:10.1007/BF01797130](https://doi.org/10.1007/BF01797130).

- de Jong, A., Alders, M., Feijten, P., Visser, P., Deerenberg, I., Van Huis, M., and Leering, D. (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, A.M., Hyndman, R.J., and Snyder, R.D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association* 106(496): 1513–1527. [doi:10.1198/jasa.2011.tm09771](https://doi.org/10.1198/jasa.2011.tm09771).
- Durbin, J. and Koopman, S.J. (2012). *Time series analysis by state space methods*. Oxford: Oxford University Press.
- Fotheringham, A.S. and O’Kelly, M.E. (1989). *Spatial interaction models: Formulations and applications*. Dordrecht: Kluwer Academic Publishers.
- Fotheringham, A.S. and Wong, D.W. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A* 23(7): 1025–1044. [doi:10.1068/a231025](https://doi.org/10.1068/a231025).
- Hamilton, J.D. (2018). Why you should never use the Hodrick-Prescott filter. *Review of Economics and Statistics* 100(5): 831–843. [doi:10.1162/rest\\_a\\_00706](https://doi.org/10.1162/rest_a_00706).
- Harvey, A.C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge: Cambridge University Press.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.
- Holt, C.C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting* 20(1): 5–10. [doi:10.1016/j.ijforecast.2003.09.015](https://doi.org/10.1016/j.ijforecast.2003.09.015).
- Husby, T.G., de Groot, H.L., Hofkes, M.W., and Dröes, M.I. (2014). Do floods have permanent effects? Evidence from the Netherlands. *Journal of Regional Science* 54(3): 355–377. [doi:10.1111/jors.12112](https://doi.org/10.1111/jors.12112).
- Hyndman, R.J. and Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. Melbourne: OTexts.
- Hyndman, R.J. and Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software* 26(3): 1–22. [doi:10.18637/jss.v027.i03](https://doi.org/10.18637/jss.v027.i03).
- Hyndman, R.J. and Koehler, A.B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting* 22(4): 679–688.

[doi:10.1016/j.ijforecast.2006.03.001](https://doi.org/10.1016/j.ijforecast.2006.03.001).

- Hyndman, R.J., Koehler, A.B., Snyder, R.D., and Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting* 18(3): 439–454. [doi:10.1016/S0169-2070\(01\)00110-8](https://doi.org/10.1016/S0169-2070(01)00110-8).
- Kalman, R.E. (1960). Contributions to the theory of optimal control. *Boletín de la Sociedad Matemática Mexicana* 5(2): 102–119.
- Kaplan, G. and Schulhofer-Wohl, S. (2017). Understanding the long-run decline in interstate migration. *International Economic Review* 58(1): 57–94. [doi:10.1111/iere.12209](https://doi.org/10.1111/iere.12209).
- Koopman, S.J. and Ooms, M. (2011). Forecasting economic time series using unobserved components time series models. In: Clements, M.P. and Hendry, D.F. (eds.). *The Oxford handbook of economic forecasting*. Oxford: Oxford University Press.
- Lee, R. and Anderson, M. (2002). Malthus in state space: Macro economic-demographic relations in English history, 1540 to 1870. *Journal of Population Economics* 15(2): 195–220. [doi:10.1007/s001480100091](https://doi.org/10.1007/s001480100091).
- Li, J.S.H., Zhou, K.Q., Zhu, X., Chan, W.S., and Chan, F.W.H. (2019). A Bayesian approach to developing a stochastic mortality model for China. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 182(4): 1523–1560. [doi:10.1111/rssa.12473](https://doi.org/10.1111/rssa.12473).
- Matthews, S.A. and Parker, D.M. (2013). Progress in spatial demography. *Demographic Research* 28(10): 271–312. [doi:10.4054/demres.2013.28.10](https://doi.org/10.4054/demres.2013.28.10).
- Monahan, J.F. (1984). A note on enforcing stationarity in autoregressive-moving average models. *Biometrika* 71(2): 403–404. [doi:10.1093/biomet/71.2.403](https://doi.org/10.1093/biomet/71.2.403).
- Mulder, C.H. (2018). Putting family centre stage: ties to nonresident family, internal migration, and immobility. *Demographic Research* 39(43): 1151–1180. [doi:10.4054/DemRes.2018.39.43](https://doi.org/10.4054/DemRes.2018.39.43).
- Ordorica-Mellado, M. and García-Guerrero, V.M. (2016). Estimating the demographic dynamic of small areas with the Kalman filter. In: Schoen, R. (ed.). *Dynamic demographic analysis*. Cham: Springer: 261–271.
- Petris, G., Petrone, S., and Campagnoli, P. (2009). *Dynamic linear models with R*. New York: Springer.
- Raftery, A.E., Li, N., Ševčíková, H., Gerland, P., and Heilig, G.K. (2012). Bayesian probabilistic population projections for all countries. *Proceedings of the National Academy of Sciences* 109(35): 13915–13921. [doi:10.1073/pnas.1211452109](https://doi.org/10.1073/pnas.1211452109).
- Raymer, J., Willekens, F., and Rogers, A. (2019). Spatial demography: A unifying

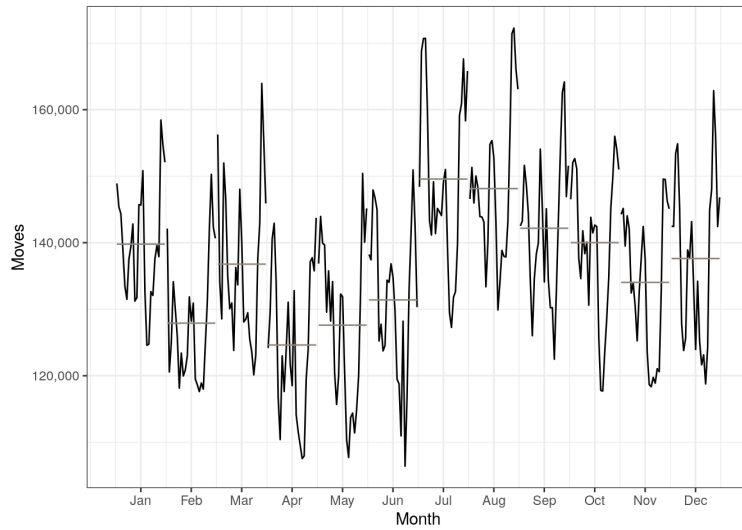
- core and agenda for further research. *Population, Space and Place* 25(4): e2179. doi:10.1002/psp.2179.
- Rueda, C. and Rodríguez, P. (2010). State space models for estimating and forecasting fertility. *International Journal of Forecasting* 26(4): 712–724. doi:10.1016/j.ijforecast.2009.09.008.
- Smith, S.K., Tayman, J., and Swanson, D.A. (2013). *A practitioner's guide to state and local population projections*. Dordrecht: Springer.
- Tashman, L.J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting* 16(4): 437–450. doi:10.1016/S0169-2070(00)00065-0.
- te Riele, S., Huisman, C., Stoeldraijer, L., de Jong, A., van Duin, C., and Husby, T. (2019). *PBL/CBS Regionale bevolkings- en huishoudensprognose 2019–2050: Belangrijkste uitkomsten*. Den Haag: Statistische Trends.
- West, M. and Harrison, J. (2006). *Bayesian forecasting and dynamic models*. New York: Springer Science and Business Media.
- Winters, P.R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science* 6(3): 324–342. doi:10.1287/mnsc.6.3.324.
- Wiśniowski, A., Smith, P.W., Bijak, J., Raymer, J., and Forster, J.J. (2015). Bayesian population forecasting: Extending the Lee-Carter method. *Demography* 52(3): 1035–1059. doi:10.1002/psp.2179.
- Young, P.C., Ng, C.N., Lane, K., and Parker, D. (1991). Recursive forecasting, smoothing and seasonal adjustment of non-stationary environmental data. *Journal of Forecasting* 10(1-2): 57–89. doi:10.1002/FOR.3980100105.
- Zietz, J. and Traian, A. (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. doi:10.1016/j.qref.2013.12.004.

## **Appendix A**

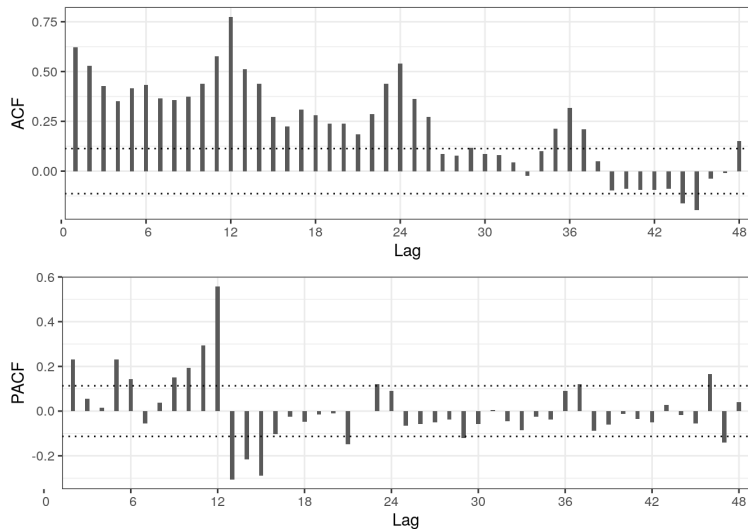
In addition to the cycle, there are strong seasonal patterns that vary over time. The seasonal subseries plot in Figure A-1 shows the movement per year and average per month (lower panel), and the horizontal lines in the figure indicating the means for each month. This plot enables us to see the underlying 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 A-2 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 ‘scalped’ 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 A-1: Seasonal subseries plot of the frequency of mobility**



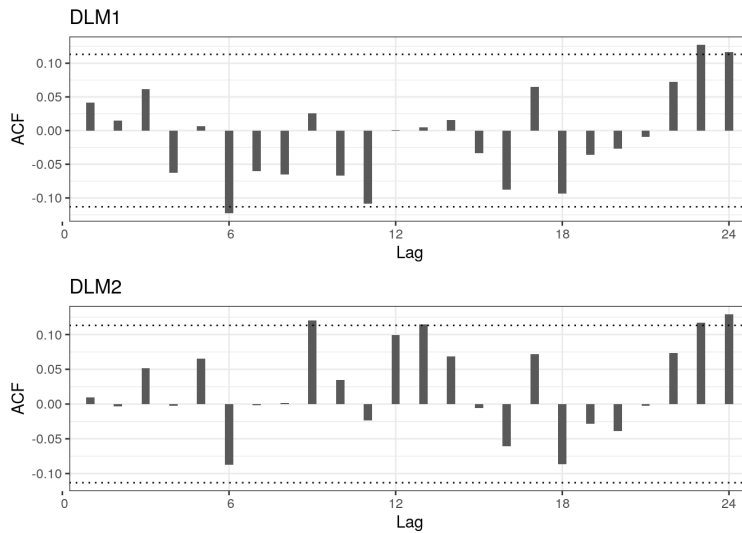
**Figure A-2: Autocorrelation function of the frequency of mobility**



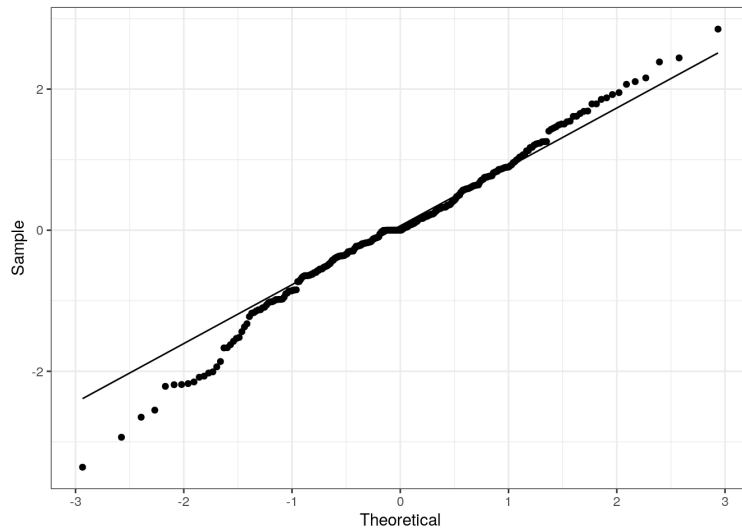
## Appendix B

This appendix presents relevant figures used for the estimation of the DLMs. The first figure shows the autocorrelation functions used for determining the order of the autoregressive elements of DLM1 and DLM2. The second figure shows whether the residuals of DLM are normally distributed.

**Figure B-1: Autocorrelation function of the standardised residuals**



**Figure B-2:** Normal probability plot of standardized one-step-ahead forecast errors of DLM2





## Appendix C

**Table C-1: MAPE of cross-validation exercise: mean, standard deviation, min and max across origins**

Horizon	Model	Mean	SD	Min	Max
6	ARIMA	0.0391	0.0108	0.0169	0.0611
	DLM1	0.0427	0.0171	0.0103	0.0689
	DLM2	0.0359	0.0105	0.0163	0.0568
	ETS	0.0366	0.0107	0.0136	0.0573
	Holt-Winters	0.0438	0.0186	0.0191	0.0743
	Naïve	0.0560	0.0216	0.0207	0.0916
	StructTS	0.0410	0.0101	0.0215	0.0620
	TBATS	0.0428	0.0081	0.0251	0.0599
12	ARIMA	0.0514	0.0168	0.0294	0.0833
	DLM1	0.0577	0.0264	0.0233	0.1029
	DLM2	0.0429	0.0096	0.0295	0.0733
	ETS	0.0453	0.0162	0.0241	0.0744
	Holt-Winters	0.0605	0.0271	0.0256	0.1051
	Naïve	0.0537	0.0143	0.0352	0.0800
	StructTS	0.0534	0.0154	0.0332	0.0774
	TBATS	0.0528	0.0158	0.0299	0.0911
18	ARIMA	0.0619	0.0266	0.0283	0.1113
	DLM1	0.0756	0.0411	0.0340	0.1524
	DLM2	0.0491	0.0096	0.0357	0.0786
	ETS	0.0526	0.0226	0.0254	0.0963
	Holt-Winters	0.0795	0.0376	0.0304	0.1473
	Naïve	0.0573	0.0149	0.0355	0.0884
	StructTS	0.0663	0.0241	0.0344	0.1060
	TBATS	0.0610	0.0277	0.0298	0.1362

