Dynamic linear models for trial heat polls in multi-party elections

Matthias Orlowski* Lukas F. Stoetzer†

First Draft: June 17, 2016
This Version: June 17, 2016

We assess the predictive quality of different dynamic Bayesian forecasting models suited for modeling the compositional data that results from trial heat polls in multi-party systems. In particular, we compare different specifications of dynamic linear models (DLM) fitted on log-ratio transformed data. In addition to existing random walk models [Jackman, 2005, Linzer, 2013, Walther, 2015], the DLM framework allows for model specifications with local trends and covariance structures that reflect the correlated evolution of party support over time. We assess the point predictions of different DLM specifications and the coverage of the corresponding predictive intervals in the run-ups to the German Federal Election of 2013. Comparing polls the DLM based forecasts, we find that the latter generate predictive intervals that more accurately represent the uncertainty about the final outcome far ahead of election day. Only within the last month prior to election day, polls offer reliable information about the expected election outcome. With these findings, we provide guidance for model selection when modeling dynamic compositional data within or beyond the specific context of elections forecasting.

*Data Scientist, SPD Parteivorstand, orlowsma@gmail.com
†Post-Doc, Department of Political Science, Massachusetts Institute of Technology, lstoetze@mit.edu

Election polls offer important information to politicians and political analysts alike, both, in evaluating ongoing political campaigns as well as forecasting upcoming elections. With polls being published on an almost daily basis, it is hardly suprizing that debates about the validity and value of trial heat polls have entered the political science literature. To what extent do polling results reveal the dynamics of campaigns and how can they be used to forecast final election results? Most contributions are focused on the U.S. where political competition boils down to the race between two major candidates [Jackman, 2014, Rigdon et al., 2009, Wlezien and Erikson, 2006, Jackman, 2005, Lock and Gelman, 2010, Linzer, 2013]. Only recently researchers turned their attention to multi-party elections which bear distinct methodological challenges [Jennings and Wlezien, 2013, Walther, 2015]. Tracing the ever-changing support of multiple parties and making sense of the correlation between weekly gains and losses presumably makes it harder to comprehend the role of polling results in the election cycle. Which also amounts to the fact that forecasting support to far-ahead elections appears far-more error prone in multi-party systems when compared to the two-party case.

This paper offers guidance how trial heat polls can be used in evaluating the ongoing support for multiple political parties, but also in forecasting the support to election day. We discuss the application of Dynamic Linear Models (DLM), specifically tailored to the situation of pooling polls for multiple parties. While a simple specification of DLM, known as the local level or random walk model, has been predominant in previous studies [Jackman, 2005, Linzer, 2013, Walther, 2015, Stoltenberg, 2013], our discussion shows that more elaborate structures might be suited to a situation with multiple parties. In particular, the evolution of support between different parties are likely to be correlated. Falling support for party A, is potentially correlated to a gain of a specific other party B. We discuss model specifications that take this intriguing fact into account. Moreover, we consider how to model upward and downward trends in a DLM framework. This extension is meaningful for parties that, over longer periods of time, constantly gain or lose popularity. Our discussion further deals with the simplex constraint inherent in compositional data of electoral support. We propose to fit models on an additive logistic transformation of predicted vote shares [Philips et al., 2016] and to then re-transform posterior and predictive distribitions to the original scale. This strategy bounds the predictive distributions to the vote share interval and thereby provides meaningful forecasts.

We exemplify our approach with an application to the Germen Federal Election campaign 2013. Our results highlight the ambiguity of poll-based predictions in forecasting eletion outcomes. Predictions based on polls early in the campaign are often to vague to be informative. DLMs offer an intriguing way to represent this uncertainty as they inherently provide a model for how the support for parties evolves over time. Although survey responses are precise in capturing the public mood at a given time-point, the state of public support is fluctuating over the course of a campaign, magnifying the uncertainty associated with predictions of an election in the distant future. The application further shows that our extensions to the independent local level models do not necessarily increase predictive power, but reveal interesting patterns of the dynamics.

1. Log-ratio transformation for poll data

Polls entail useful information to trace a party's support in an ongoing election campaign and to forecast expected electoral support to election day. Research that relies on this information commonly models a vector of polling information of proportional support for each party over time [Walther, 2015, Jennings and Wlezien, 2013, Jackman, 2005, Jennings and Wlezien, 2013]. However, the proportions are subject to an important constraint: the sum of all proportions is one, forcing each share to be bound to the unit interval.[1] The data structure can be defined as follows. $\boldsymbol{Y}_t$ is a $K \times 1$ vector of compositional data with elements y_{1t}, such that $\sum_k^K y_{kt} = 1$ and $0 \leq y_{kt} \leq 1 \; \forall k \in \{1, \ldots, K\}$. $t = \{1, \ldots, T\}$ indexes time and $k = \{1, \ldots, K\}$ categories.

To integrate the simplex constraint in the study of trial heat polls, we propose to model log ratios of shares instead of shares themselves. Philips et al. [2016] recently suggest to use the log-ratio transformation for dynamic compositional data [see also Tomz et al., 2002, Katz and King, 1999]. This suggestion goes back to Aitchison's work on models for compositional data [Aitchison, 1982, 1986].[2] For K parties, where category K is the reference-party, the transformation gives a $K-1$ vector:

$$\boldsymbol{Y}_t^* = \left[ln \left(\frac{y_{1t}}{y_{Kt}} \right), \ldots, ln \left(\frac{y_{K-1t}}{y_{Kt}} \right) \right]. \tag{1}$$

This changes the interpretation of the margin of error generally associated with the polling means. The error variance of proportion is calculated by $\sigma^2_{y_{kt}} = \frac{\boldsymbol{y_{kt}}(1-\boldsymbol{y_{kt}})}{N_t}$, where N_t is the number of survey respondents. Researcher use this equation to calculate the error variance beforehand and integrate the information in their models[3]. Employing log-odds changes the calculation of th error variance for the survey instrument.[4] Approximating the variance-covariance matrix using first-order Taylor series expansion gives:

[1] Most applications ignore this issue and directly model the vector of shares. While for two party systems (with expected vote shares around fifty percent) this might be sensible, for systems with multiple parties most often generates implausible and undesired forecasts.

[2] Aitchison [1982] also highlights the suboptimal use of dirichlet class distributions in modeling compositional data. "A major obstacle to its [Dirichlet distribution] use in the statistical analysis of compositional data is that it seldom, if ever, provides an adequate description of actual patterns of variability of compositions" [Aitchison, 1982, p.142]. First, the probability contours of the Dirichlet distribution are convex, failing to describe concave patterns accurately. Second, the dirichlet distribution imposes a strong independence structure that is rarely meet. We follow this advice and focus on transformation of shares instead of applying dynamic dirichlet models to the data-structure [Da-Silva and Rodrigues, 2015]

[3] In the case of multiple shares there are also covariances in measurement errors to be considered. The covariances are $\frac{-y_{kt}y_{qt}}{N_t}$, where $j \neq q \in \{1, \ldots, K\}$. For a more detailed discussion please refer to the appendix A

[4] The derivation of the moments are presented in the appendix A

$$\boldsymbol{V}^*_{\boldsymbol{Y_t}} = \begin{pmatrix} \frac{1}{N_t}\left[\frac{1-y_{1t}}{y_{1t}} + \frac{1-y_{Kt}}{y_{Kt}} + 2\right] & \cdots & \frac{1}{N_t}\left[1 + \frac{1-y_{Kt}}{y_{Kt}}\right] \\ \vdots & \ddots & \vdots \\ \frac{1}{N_t}\left[1 + \frac{1-y_{Kt}}{y_{Kt}}\right] & \cdots & \frac{1}{N_t}\left[\frac{1-y_{K-1t}}{y_{K-1t}} + \frac{1-y_{Kt}}{y_{Kt}} + 2\right] \end{pmatrix}. \tag{2}$$

The equation can be deployed to model evolving support of the log-odds over time. As shall be discussed below, calculating the observational variance is particular meaningful in the application of dynamic linear models to trial heat polls.

2. Specification of dynamic linear models for trial heat polls

Dynamic linear models are formulated by two equations, the observation- and the state-equation [West and Harrison, 1997, Petris et al., 2008]. In the most general case the observation equation is given by:

$$\boldsymbol{Y_t} = \boldsymbol{F}_t\boldsymbol{\theta}'_{\boldsymbol{t}} + \boldsymbol{\nu_t}, \quad where \quad \boldsymbol{\nu_t} \sim N(\boldsymbol{0}, \boldsymbol{V_t}). \tag{3}$$

where $\boldsymbol{Y_t}$ are the observed outcomes. The formulation allows for univariate, but also multivariate time lines. $\boldsymbol{\theta}_t$ is a vector of state parameters that, for example, contain the level-support for a particular party. $\boldsymbol{F_t}$ is a design matrix that relates states to observed outcomes. $\boldsymbol{\nu_t}$ is the observational error that are assumed to be multivariate normal distributed with variance-covariance matrix $\boldsymbol{V_t}$. Additionally, each DLM is specified by a system, or state equation:

$$\boldsymbol{\theta_t} = \boldsymbol{G}_t\boldsymbol{\theta}'_{\boldsymbol{t-1}} + \boldsymbol{\omega_t}, \quad where \quad \boldsymbol{\omega_t} \sim N(\boldsymbol{0}, \boldsymbol{W_t}). \tag{4}$$

$\boldsymbol{G_t}$ is a design matrix that relates prior states to present states. $\boldsymbol{\omega_t}$ is the evolution error of the sates, assumed to be multivariate normal distributed with variance-covariance matrix $\boldsymbol{W_t}$. $\boldsymbol{W_t}$ then depicts the volatility of the system over time. Each DLM is defined by a quadruple $(\boldsymbol{F_t}, \boldsymbol{G_t}, \boldsymbol{V_t}, \boldsymbol{W_t})$.

In our application of DLMs to trial heat polls, the observed values are the log-ratio transformed vector of proportions $\boldsymbol{Y_t^*}$ and $\boldsymbol{V_t}$ is defined by the margin-of-error of the transformed data $\boldsymbol{V^*_{Y_t}}$ (see equation 2). If we are furthermore willing to impose a time-invariant structure, a model specification only requires choosing an appropriate design matrices $(\boldsymbol{F}, \boldsymbol{G})$ for a vector of state parameter $\boldsymbol{\theta_t}$, and a variance-covariance structure of how the states evolve over time $\boldsymbol{W}$. The next subsections outlines three potential specifications.

2.1. A local level model

A commonly used model in the literature is the so-called local level or random walk model [see e.g. Jackman, 2014, Walther, 2015, Linzer, 2013]. In this the state parameter are defined by the level of support for a party at a given time point $\boldsymbol{\theta}_t = \boldsymbol{\mu}_t$, where $\boldsymbol{\mu}_t$ is row vector of size $K-1$. Most existing models assume that the evolution of one party's support is not linked to the evolution of any other party's support. Thus, the design matrices $\boldsymbol{G}$ as well as $\boldsymbol{F}$ are set equal to an identity matrix $\boldsymbol{I}$ of size $[K-1 \times K-1]$. The observed values for a party only depend on the level support for this party. This results in a observable equation and state equation of:

$$\begin{aligned} \boldsymbol{Y}_t &= \mathbf{I}\boldsymbol{\mu'}_t + \boldsymbol{\nu}_t, \quad \textit{where} \quad \boldsymbol{\nu_t} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{V_{y_t}}), && (5) \\ \boldsymbol{\mu}_t &= \mathbf{I}\boldsymbol{\mu'}_{t-1} + \boldsymbol{\omega}_t, \quad \textit{where} \quad \boldsymbol{\omega_t} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{W_\mu}). && (6) \end{aligned}$$

The level of support at time point t for party k $(\boldsymbol{\mu_t})$ systematically only depends on the previous level for this party. Independence of the evolution structure further implies that the state variance are uncorrelated across different levels of party support. Hence, $\boldsymbol{W_\mu} = diag(\boldsymbol{\sigma_\mu})$, where $\boldsymbol{W_\mu}$ only has entries on the diagonal, $\boldsymbol{\sigma_\mu} = [\sigma_1^\mu, \ldots, \sigma_{K-1}^\mu]$. If researcher moreover ignore the covariance in observational errors $\boldsymbol{V_{y_t}}$, the simplest multivariate local level model could be estimated independent for each party[5]. In this specification observed values randomly evolve over time. Next, we discuss an additional aspect of the evaluation structure.

2.2. A local trend model

Additional to the level of support, trial heat polls suggest certain trends. Parties often seem to be in an upward or downward trend that appears president over weeks, or even month. New arriving poll results are for weeks systematically lower or higher as the previous level would suggest. A local trend model includes such trends in the DLM formulation. Additional to the support level, the state parameters contain a local trend parameter $\boldsymbol{\alpha_t}$ for the $K-1$ parties $\boldsymbol{\theta_t} = (\boldsymbol{\mu_t}, \boldsymbol{\alpha_t})$. The complete model is written out

$$\boldsymbol{Y}_t^* = \boldsymbol{F\theta'}_t + \boldsymbol{\nu}_t, \quad where \quad \boldsymbol{\nu_t} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{V_{y_t^*}}), \tag{7}$$

$$\boldsymbol{\theta}_t = \boldsymbol{G\theta'}_{t-1} + \boldsymbol{\omega}_t, \quad where \quad \boldsymbol{\omega_t} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{W_\theta}). \tag{8}$$

Where $\boldsymbol{F}$ is a design matrix of size $[2*(K-1) \times 2*(K-1)]$ and $\boldsymbol{F} = \left(\begin{smallmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{smallmatrix}\right)$, such that only the levels of party-specific support appear in the observational equation. For the state equation, the current support level is affected by the party-specific previous levels and the previous trends, capturing the idea that a party can generally be in an upward or downward period. The momentousness trend is a random walk that originates from

[5] In fact, this is the most popular model specification for untransformed proportions: $y_{tk} = \mu_{tk} + \nu_{tk}$ and $\mu_{tk} = \mu_{t-1k} + \omega_{tk}$, where $\nu_{tk} \sim N\left(0, \frac{\boldsymbol{y_{kt}}(1-\boldsymbol{y_{kt}})}{N_t}\right)$ and $\omega_{tk} \sim N(0, \sigma_k^\mu)$.

of prior trend. This idea is implemented by a matrix $\boldsymbol{G} = \left(\begin{smallmatrix} \boldsymbol{I} & \boldsymbol{I} \\ \boldsymbol{0} & \boldsymbol{I} \end{smallmatrix}\right)$.

In this specification, the evolution variance again is assumed to be uncorrelated across parties.

$$\boldsymbol{W_\theta} = \begin{pmatrix} \boldsymbol{W_\mu} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{W_\alpha} \end{pmatrix} \tag{9}$$

where $\boldsymbol{W_\mu} = diag(\boldsymbol{\sigma_\mu})$ is the same as in the local trend model. For the $\boldsymbol{W_\alpha} = diag(\boldsymbol{\sigma_\alpha})$, where $\boldsymbol{\sigma_\alpha} = [\sigma_1^\alpha, \ldots, \sigma_{K-1}^\alpha]$. While the local trend model potentially accounts for an interesting aspect of polls it still imposes the strong assumption that trends and levels for parties evolve independent.

2.3. Covariance in state evolution

The independent evolution of states in the described models can be relaxed by introducing meaningful covariance structures in $\boldsymbol{W_\theta}$. Covariances can be integrated for the evolution of local trends as well as local levels. For the local level model, the covariance terms measure how strongly the volatility of levels are correlated between parties. One party's positive random shifts are negatively correlated with another party's shift. Instead of setting all off-diagonal elements in W_μ to zero, the covariances in level evolution can be specified and estimated:

$$\boldsymbol{W_\mu} = \begin{pmatrix} \sigma_1^\mu & \sigma_{12}^\mu & \cdots & \sigma_{1K-1}^\mu \\ \sigma_{12}^\mu & \sigma_2^\mu & \cdots & \sigma_{2K-1}^\mu \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1K-1}^\mu & \sigma_{2K-1}^\mu & \cdots & \sigma_{K-1}^\mu \end{pmatrix} \tag{10}$$

Covariance in state evolution might also be meaningful in the of case local trends. Naturally, if one party experiences a strong upward trend, others parties should be loosing votes. Covariances might be able to pick this up. The basis for this model specification is the local trend model specified above (equation 7), but with covariance between the trends of the parties, resulting in:

$$\boldsymbol{W_\alpha} = \begin{pmatrix} \sigma_1^\alpha & \sigma_{12}^\alpha & \cdots & \sigma_{1K-1}^\alpha \\ \sigma_{12}^\alpha & \sigma_2^\alpha & \cdots & \sigma_{2K-1}^\alpha \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1K-1}^\alpha & \sigma_{2K-1}^\alpha & \cdots & \sigma_{K-1}^\alpha \end{pmatrix} \tag{11}$$

In total the description creates three different model specifications: 1) an independent local level, 2) a local level model with covariance in evolution and 3) a local trend model that allows for covariances. The next section describes our Bayesian estimation strategy for the different models.

3. Bayesian inference for Dynamic Linear Models

Methods for Bayesian estimation of DLM are well-established [Frühwirth-Schnatter, 1994, Carter and Kohn, 1994, Shephard, 1994] and are discussed in seminal volumes on the subject [West and Harrison, 1997, Petris et al., 2008].[6] In the application of DLM to trial heat polls, we are interested in the estimation of two sets of parameters: the states $\boldsymbol{\theta_t}$, and the evolution variance $\boldsymbol{W_\theta}$. Prior knowledge about those are expressed through a probability distribution about the evolution variance $P(\boldsymbol{W_\theta})$ and the inital state $P(\boldsymbol{\theta_0})$. Given this prior beliefs and the assumptions that the state space is a Markov chain and conditionally on the states, the observed values are independent and depend on the state only, the posteriori distribution can be expressed as:

[6]Alternatively, estimates for the states and model parameter can be obtained by maximizing the normal likelihood distribution [Petris et al., 2008, chapter 4.1].

$$P(\boldsymbol{\theta_{0:T}}, \boldsymbol{W_\theta} \mid \boldsymbol{Y}^*_{1:T}) \propto P(\boldsymbol{W_\theta})P(\boldsymbol{\theta_0}) \prod_{t=1}^{T} P(\boldsymbol{Y}^*_{1:T} \mid \boldsymbol{\theta_t}, \boldsymbol{W_\theta})P(\boldsymbol{\theta_t} \mid \ \boldsymbol{\theta_{t-1}}, \boldsymbol{W_\theta}). \tag{12}$$

A common approach to approximate this posteriori distribution is a forward filtering, backward sampling within an Gibbs sampler algorithm[Frühwirth-Schnatter, 1994]. First, the states are sampled from the full conditional distribution $P(\boldsymbol{\theta_{0:T}} \mid \boldsymbol{W_\theta}, \boldsymbol{Y^*_{1:t}})$ by running a Kalman filter [Kalman, 1960] over the time line and then using backward sampling to obtain draws for each state. Second, the parameters (here the evolution variance) are sampled given the states $P(\boldsymbol{W_\theta} \mid \boldsymbol{\theta_{0:T}}, \boldsymbol{Y^*_{1:t}})$. In the case of a conjugate prior distribution for the the evolution variance this constitutes a simple Gibbs step in the algorithm. For a more detailed description of this algorithm, the conjugate prior specifications and software implementation please refer to the appendix B. It shall be mentioned that the algorithm naturally takes care of missing-data by extrapolating missing states from prior states during the filtering step. This means that it can be straightforwardly applied to daily evolving campaigns, even though poll results might not arrive on a daily basis.

3.1. Forecasting support on election day

One argument for the application of Bayesian DLM's in trial heat polls is the ability to forecast support to election day from any point in time, and (more important) to clearly pin-down the precision of those forecasts. While for other approaches it is debatable how to correcly asses uncertainties, the dynamic framework results in a intuitive interpretation of how uncertainty evolves over time. In essence, the state evolution variances contain information about how stable (or volatile) the system will behave in the future. In case of a local level model, for example, for teh expected support on the next day ahead t she support changes according to the evolution variance. Two days make twice the variance, three days result in three times the variance and so on. Hence, the uncertainty to far ahead time points increases because the system is evolving over time.

This intuitive idea directly emerges from the DLM specification. Suppose that data is available up to time t. The k-step ahead predictive distribution for the states is simply obtained by integrating out the prior state:

$$P(\boldsymbol{\theta_{t+k}} \mid \boldsymbol{Y^*_{1:t}}) = \int P(\boldsymbol{\theta_{t+k}} \mid \boldsymbol{\theta_{t+k-1}}) P(\boldsymbol{\theta_{t+k-1}} \mid \boldsymbol{Y^*_{1:t}}) d\boldsymbol{\theta_{t+k-1}}. \tag{13}$$

The primary interest are the expected support and the variance of this distribution on election day. For the outlined general DLM (3 and 4) the integral can be computed explicitly resulting in a Gaussian predictive distribution [see e.g. Petris et al., 2008, p.71]. The time-invariant structure we impose for the specific applications ($\boldsymbol{G}, \boldsymbol{F}$ and $\boldsymbol{W_\theta}$) further simplifies the general expressions. Based on the current state, with mean vector $\boldsymbol{a_t}(0) = \boldsymbol{m_t}$ and variance-covariance $\boldsymbol{R_t}(0) = \boldsymbol{C_t}$, the expectation and variance can be calculated sequentially:

$$E\left(\boldsymbol{\theta_{t+k}} \mid \boldsymbol{Y^*_{1:t}}\right) \quad = \boldsymbol{a_t}(k) \quad = \boldsymbol{G}\boldsymbol{a_t}(k-1) \tag{14}$$

$$Var\left(\boldsymbol{\theta_{t+k}} \mid \boldsymbol{Y^*_{1:t}}\right) \quad = \boldsymbol{R_t}(k) \quad = \boldsymbol{G}\boldsymbol{R_t}(k-1)\boldsymbol{G'} + \boldsymbol{W_\theta} \tag{15}$$

For the *local-level model*, where $\boldsymbol{G} = \boldsymbol{I}$ and $\boldsymbol{\theta_t} = \boldsymbol{\mu_t}$, the predictive distribution has expectation that is equal to the current state $E\left(\boldsymbol{\mu_{t+k}} \mid \boldsymbol{Y^*_{1:t}}\right) = \boldsymbol{m_t}$, with increasing variance $Var\left(\boldsymbol{\mu_{t+k}} \mid \boldsymbol{Y^*_{1:t}}\right) = \boldsymbol{C_t} + k * \boldsymbol{W_\theta}$. For the *local-trend model*, the expectation further depends on the current trends. Although these trends are important for the predictive distribution, the main interest lays in the predicted support. Splitting up state means in levels and the trends ($\boldsymbol{m_t} = \{\boldsymbol{m_t^\mu}, \boldsymbol{m_t^\alpha}\}$) results in the expectation: $E\left(\boldsymbol{\mu_{t+k}} \mid \boldsymbol{Y^*_{1:t}}\right) = \boldsymbol{m_t^\mu} + k * \boldsymbol{m_t^\alpha}$. This implies that the local-trends are added to the expectation for every step ahead. A party in an upward trend is expected to gain more support as the current level suggest, and vise versa. Analogously, the predicted variance for the levels also depends on the variability of the trends. *Covariances in state evolution* generelly do not influence the expectations, but only impact the variance-covariance of the predictive distribution. In howfar this aspect is crucial to the analysis of trial heat polls, might best be seen by the means of a concrete example.

4. Application to the German Federal Election 2013

The German Federal Election of 2013 took place on 22th of September. The coalition of the Christian Democratic Union (CDU) and the Christian Social Union won nearly 42% of the votes. The Social Democratic Party (SPD) received 29.4 %, the Left (DIE LINKE) 8.2 % and the Alliance '90/The Greens (GRÜNE) 7.3 %. The Free Democratic Party (FDP), former coalition partner of the CDU/CSU, fall short of the 5% percent hurdle, with 2.4 % and consequently was not able to secure any seats. The result called for a great-coalition between CDU/CSU and the SPD that guaranteed Angela Merkel another term as chancellor.

In order to trace the support for the parties since the last election (27th September 2009), we combine information from several German pollsters. German's major poll agencies are Allensbach, Emnid, Forsa, Forschungsgruppe Wahlen, GMS Dr. Jung GmbH, Infratest dimap, and Institut für neue soziale Antworten (INSA). We use all polls conducted by these agencies after the last election. Our analysis concentrates on the five major parties, combing support for the CDU and CSU and collapsing support for other parties into a six category[7].

[7]The largest parties in this category are the Pirate Party (PIRATEN) and Alternative for Germany (AFD). The Pirate Party received considerable support (up to 10%) during 2011 and 2012, but finished with 2.2 %. The AFD entered the political sphere beginning of 2013 and close to the election polled around 4% points

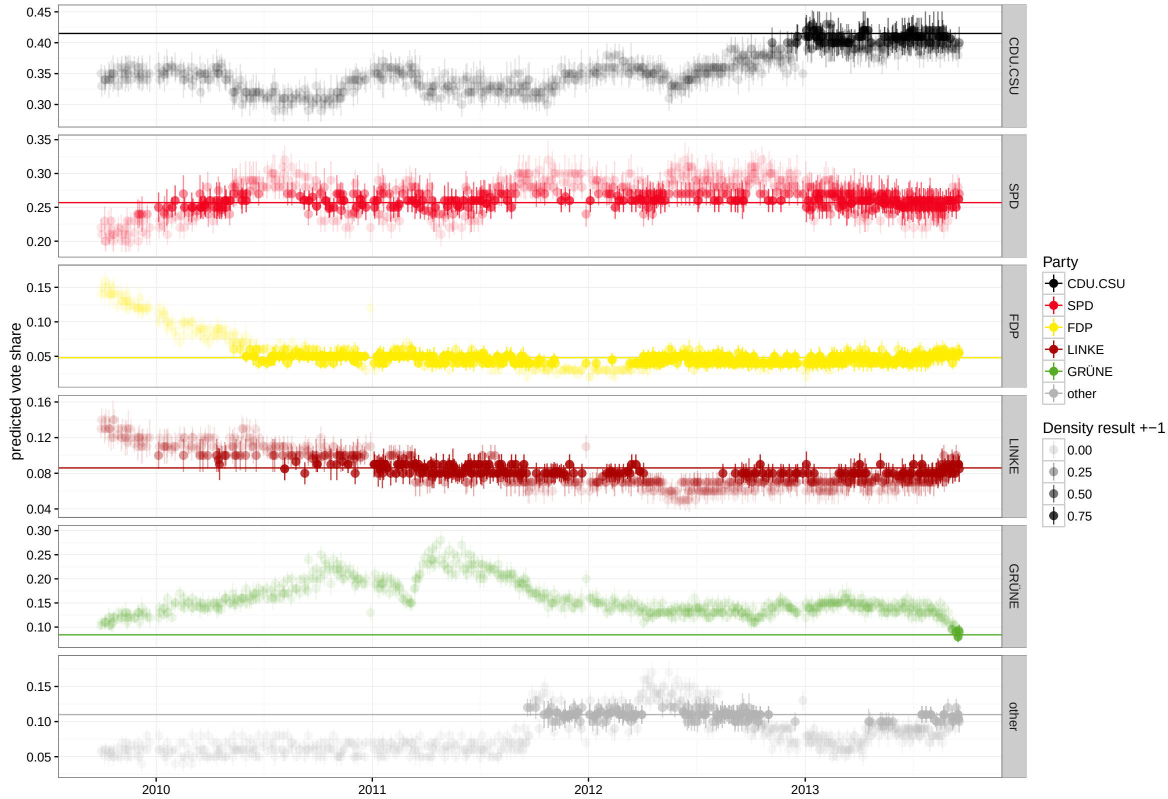

Figure 1: Poll results over the 2013 Federal Election campaign. The dots indicate average support calculated from one particular poll. Confidence Intervals are calculated based on the margin of error. The density of the dots and intervals depicts the coverage of plus/minus one percent of the final result.

Figure 1 shows the daily poll results for the the time between the 2009 election and the 2013 election. The dots indicate the the average support in one particular poll. The confidence intervals are calculated using the margin of error. One first point to note is that those intervals are generally relatively narrow, as the margin of error is small with surveys that poll around 1000 people. E.g. for the SPD who polls around the 25 percentage line the margin of error is around 0.01. The 95% confidence interval covers plus or minus 2 percentage points. Those interval have to be carefully interpreted. They do not at all imply that the final result will be found within this interval. They rather provide uncertainty about a snapshot of the mood in a given time point from a frequentest perspective. If the poll on the same day would be administered over and over again, roughly 95% of the means would be found in this interval. It is entirely incorrect to suggest that the final result will be found within the interval. In fact, the evolution of support suggest that such uncertainty calculations cannot account for the strong variation that the time-line actually experiences.

Instead, the graph underscores that the intervals most often do not include the final result. To see this more clearly, we calculate a measurement of predictive accuracy. The measurement takes two aspects into account. First, how strongly the average deviates from the target. Second, how strongly the uncertainty distribution around this estimate overlaps with the final result. The second aspect is important in our later analysis, as wider predictive distributions might be off on average, but still include the final result. This desirable aspect of a predictive distribution, however, can have different degrees. An interval that spans the complete result space (in our case always ranges from 0 to 1) always includes the final result, but as such is less desirable as a distribution interval that contains the final result and covers a smaller range. To consider this aspect we employ the density of a distribution around the final result. We calculate how much density of the normal distributed errors falls within the interval plus/minus one percentage of the final result. The larger this density, the more accurate the prediction.

Shading the intervals with our density measurement of predictive accuracy in Figure 1 highlights how well the polls relate to the election result. For example, for the CDU/CSU the polls in the first three years are far below the final result and hence have a small density in predicting the outcome. Only within the last year the polls generate some density around the target. The increasing coverage over time is particular striking for the GRÜNE who in the mid-off 2011 polled around 20% which is far above their final 7.2%. Only within the last weeks the polls accurately cover the target. The FDP started off from their strongest election result with 14.6% in 2009, but the polls quickly faded to the five percent level, with more-or-less accurately predicted the final outcome.

What is there to learn from this? One the one hand, it is uncontroversial to state that polls administered years before the election are ill-suited to directly infer about the support on election day. They rather provide a snapshot of the current support for parties. On the other hand, it should be clear that the time-lines convey information that can be put to use in forecasting the final outcome. We are of the conviction that the DLM framework can help to form reliable and accurate forecasts. Especially, since the formulations concretize ideas how the support evolves over time. This is a central aspect in understanding what the support will be on election day and how certain one about the forecasts. It remains undisclosed which specification most reliably mimics the evolution. Do trends matter? For some of the time-lines it seems that way. For example, the green party was on a downward trend since mid 2011 and the CDU experienced a strong upward trend starting one and a half years before the election. Do covariances in evolution matter? This is harder to spot from the dynamics in the figure and better assets by estimating the models based on the data. We asses the predictive qualities of our DLM specifications in the next section.

4.1. Results from Dynamic Linear Models

This section describes the estimates obtained from three different DLM specifications for daily poll data: a local level model with all covariances set to zero; a local level model with specified evolution covariances and a local trend model with covariances on both the levels and the trends[8]. For the transformation of the proportions to log-ratios, the CDU (as the largest party) is used as the reference category[9]. We express our prior beliefs about the initial state and the evolution variances in terms of the distributions outlined in the discussion on the sampling scheme in appendix B. For the initial states

[8] Whenever there is available data from multiple polls for one day, we form a weighted averages between the results (using the number of respondents) and adjust the margin of error.

[9] The first-order taylor approximation of the moments (see appendix A) works best with a large reference category

we choose informative priors centered around the election results of 2009, with a standard deviation that corespondents to the comprehensive survey of all voters in the last election[10]. Expressing honest prior beliefs about the evolution variance is more difficult, as both the expectations and the precisions of log-ratio transformed evolution variance have to be preassigned. We generally rely on uninformative distributions that adhere to the idea that we do not know (per se) how the states will evolve, but center them around the observed covariances-variances[11]. The FFBS Gibbs sampling algorithm is run for a large amount of iterations to ensure convergence of the chain (50'000, discarding the first 30'000 as burn-in and saving every 10th iteration).

We first turn our attention to the estimates of the level support $\boldsymbol{\mu_{1:T}}$. Figure 2 presents the evolution of support (mean and 95% credible interval) for the parties from the different models[12]. It is difficult to spot any major differences between the evolution of party support. All time-lines are almost identical in that they closely follow the observed poll results. This is not surprising, as the major factor that determines how strongly the levels are updated, the signal-to-noise ratio, is similar in all of them. The signal-to-noise ratio is generally defined as the ratio of the state evolution variance compared to the observational variance $r = \frac{\boldsymbol{W_\theta}}{\boldsymbol{V_{y_t}}}$. In our models we predefine the observational variances, hence the leftover variation influences the estimation of the evolution variances, resulting in nearly equal ratios for the levels. Because of this, we observe similar adjustments to newly arriving data points. Take for example the outlier result for FDP at turn of the

[10]Specifically, our priors on the initial state is the log-ratio transformed result from the 2009 general election, where the CDU/CSU obtained 0.338, the SPD 0.23, the FDP 0.146, Die Linke 0.119, and the Greens 0.107 of the total votes cast, while 0.06 where obtained by smaller other parties. Based on these results, we also computed the prior on the initial state evolution variance. We computed the variance-covariance matrix of the log-ratio transformed proportions using the approximation described in the appendix and divided by the total votes cast in 2009, 43.371.190. This indicates our belief that the 2009 election result revealed the true state of party support at that point in time.

[11]For the Wishart distribution that describes our prior on the precision of the state evolution, we chose only one degree of freedom to express this uncertainty. We set the corresponding Σ to a diagonal matrix with the mean variance of log-ratio transformed poll results in our data (0.083). Similarly, for the local trend model (without covariances), we set the expected value of an weekly informative prior to the mean variance in changes in log-ratio transformed polls (0.027).

[12]For the presentation we re-transformed the alr-shares to the original scale.

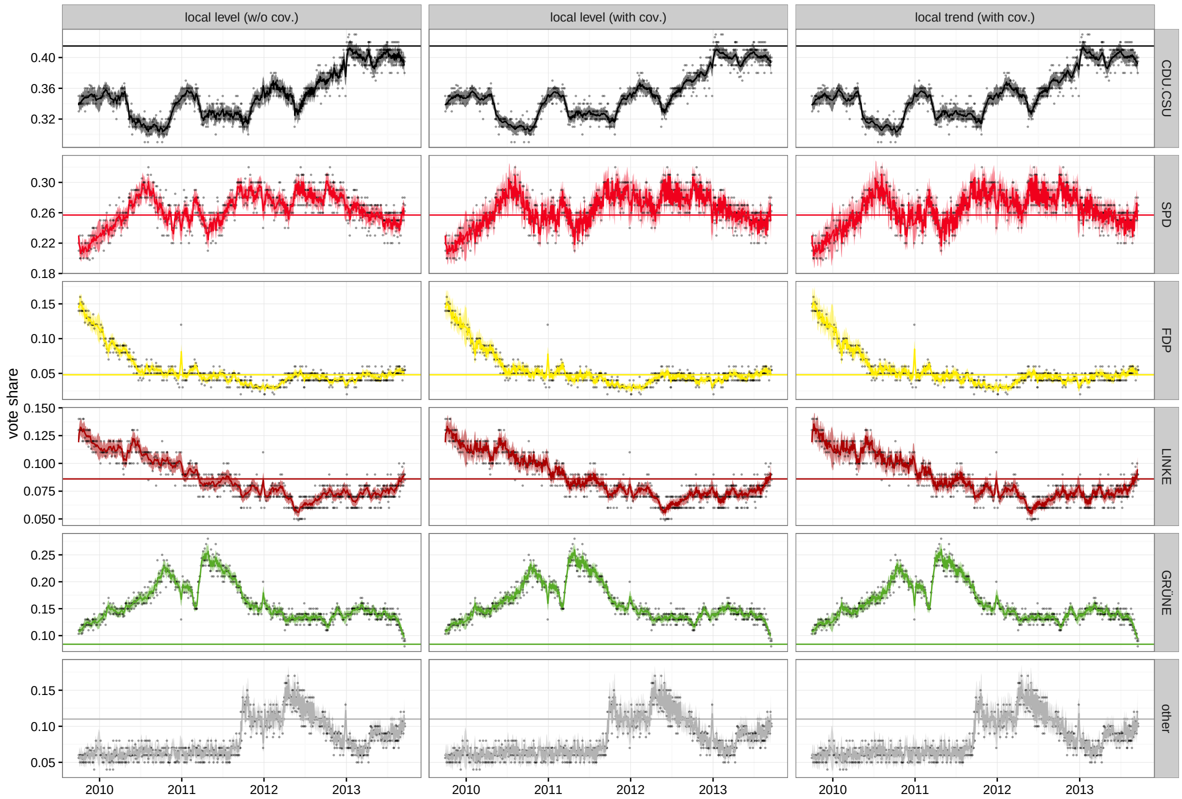

Figure 2: Level support for parties over time obtained from three different model specifications. A local level model without covariances; a local level model with covariances and a local trend model with covariances. For the presentation the log-ratio transformed shares are re-transformed to the original proportions.

year 2011. The spike that results from this is nearly identical for all models. In this extreme case, the adoption is not strong enough to include the misguiding poll within the confidence bounds. Overall, however, the states strongly adjust to the observed polls, which is again is best explained in terms of the signal-to-noise ratio. Because the margin of errors associated with the polls is small, this ratio is large, implying that the new observations are taken as informative signals about the state of support. There are nuanced differences in the uncertainties associated with the level support. For example, the credible intervals are larger for the SPD-levels when including covariances in evolution, but are smaller for the CDU-levels.

The other estimates obtained from the models, the evolution variances $\boldsymbol{W_\theta}$ and the local trends $\boldsymbol{\alpha}_{1:T}$, reveal distinct aspects about the dynamics, but do not affect the general tendency to closely mimic the poll support-patterns. Moving from an independent local level model to a local level model that includes the covariances in state evolution, uncovers interesting correlation between the random shocks in support (see appendix D). For example, an increase in support for the SPD (compared to the CDU) is correlated with an increase of the FDP support (compared to the CDU). Such correlations relate to concrete research questions about voter transitions. In this way, covariance specifications might permit researcher to make inference about the connectives of parties' platforms, and/or voters' preference orderings. The trends itself are of particular concern to political analyst who want to judge whether an increase in support can be labeled as a trend. However, for the application at hand it is challenging to identify any major trends (see appendix E). Only for the GRÜNE (compared to the CDU) there appear to be some signifcant upward (beginning of 2011), but also downward (shortly before the election) trends.

Based on the estimates alone it is hard to decide which of the model specifications is most appropriate for our application. While aspects of the extensions might be insightful in their own right, the evolution of support levels exhibit only minor differences. This does not mean that there no pronounced disparities in forecasting support to election day. The next section, therefore, discusses how well those different model specifications predict the final outcome.

4.2. Comparison of predictive distributions

In order to asses the predictive qualities of the models, we make forecasts from one day, one week, one month and one year before the election[13]. Therefore, we run all models on the full data-set and three sub-sets of the time-line. The subsets have different cut-off

[13]We hope to increase this analyses to more time-points in the near future

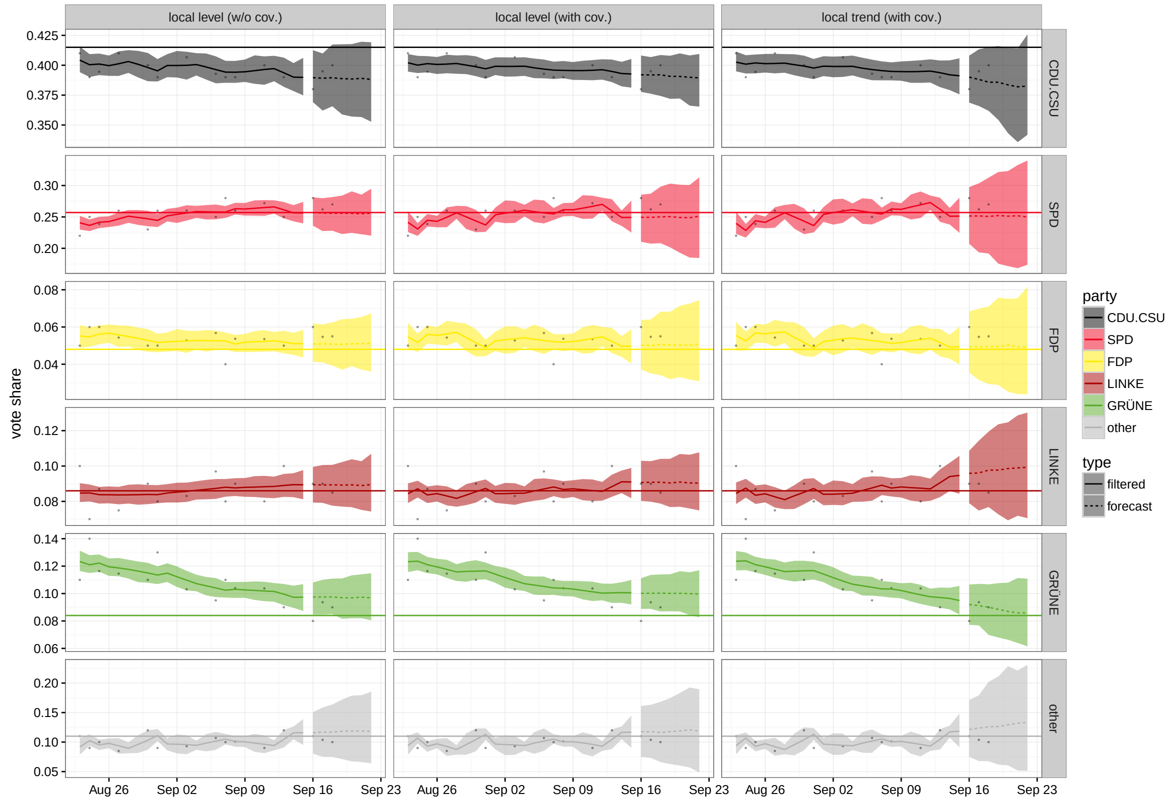

Figure 3: Predictive intervals foe one week forecast.

points. The one-week forecast only uses information up to one week before the election, the one-year subset is based on polls administered up to one year prior to the election, and so on. This approach is superior, as we should not use information that wouldn't have been available at this time[14].

The way how the predictive distribution develops is illustrated in Figure 4. It shows the filtered levels up to one week prior to the election, and the forecast distribution from this point on. In line with the theoretical description, the predictive interval gets larger with each extrapolated day. How strongly this is the case, slightly depends on the model specification and the party under investigation. Almost all predictive intervals include the final result (expect for the local level with cov. for the CDU). In order to achieve

[14] Including all time-points to estimate the model and still forecast from different days prior to the election, would influence the estimates of the evolution variances and thereby the predictive distribution.

model	one year	one month	one week	one day
survey	0.039	0.026	0.013	0.010
local level (w/o cov.)	0.063	0.022	0.013	0.011
local level (with cov.)	0.075	0.022	0.013	0.012
local trend (with cov.)	0.186	0.027	0.017	0.013

Table 1: Root mean squared forecasting error over all parties by model and forecasting date.

this, however, the one week predictive intervals already span a wide range. For example, the SPD predictive interval ranges from 19% to over 30%, which does not exactly results in a precise forecast. Another aspect to note is the distinction between the model that allows for trends and the level-only-models. For the latter the mean forecasts are equal to the last mean level. For the trend model, the mean forecasts can generally increase or decrease. For the GRÜNE (who are in a downward trend the week before the election), the mean of the predictive distribution goes down, producing a more accurate forecast to election day compared to the other models. Thus, with a clear trend (that further stay steady), this model should be better in foretelling the final outcome.

The first property by which to asses the forecasts are the root mean square errors for the different models. Table 1 displays the root mean square errors in comparison to the forecasts from the survey. For the survey we use the proportions calculated from one poll, and for the models the mean of the predictive distribution. Surprisingly, for the one year ahead forecast the survey indicates the smallest root mean square error. The poll that generates this forecast was an outlier at this early time point. It can be identified in Figure 1 as the one data point for the CDU (around Sep 2012) with an interval that gets close the final result. The local level model down weight this observation in comparison to the preceding levels and consequently do not preform as accurately. The local level models preform better compared to the local trend model. For the one month forecast, the local level model are most accurate, followed by the

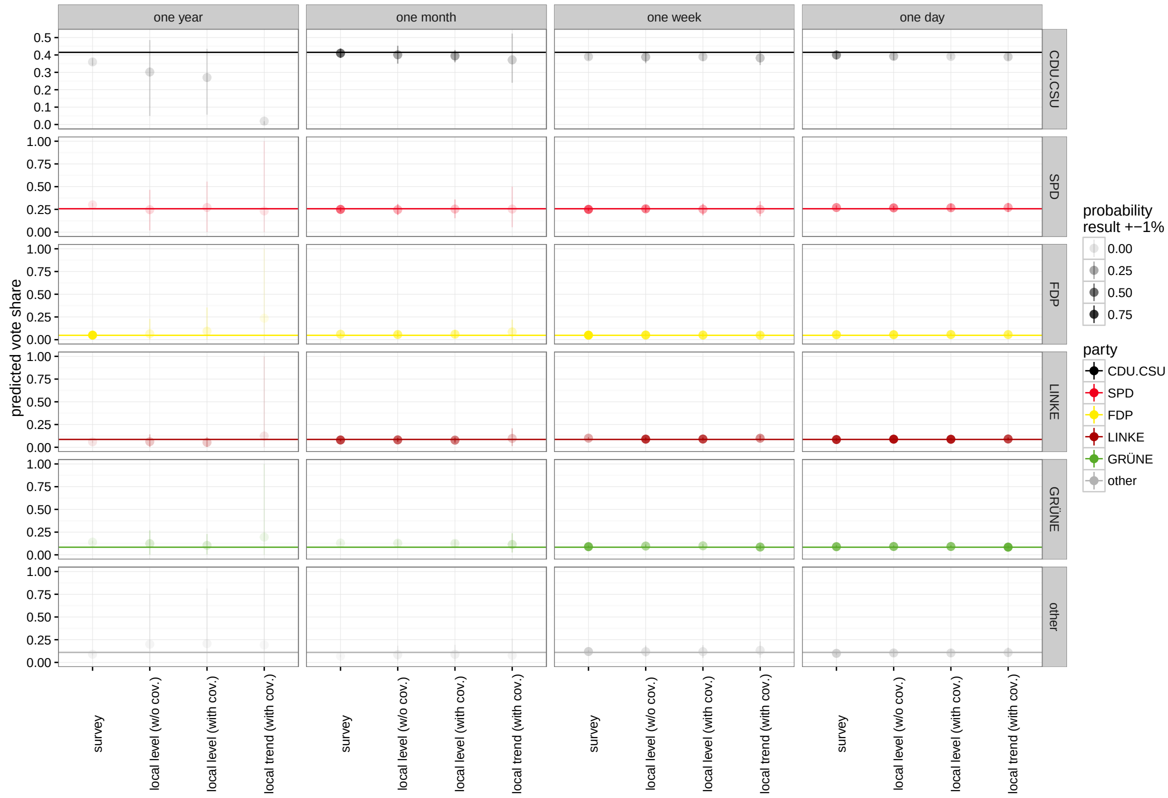

Figure 4: Predictive Intervals from different models for election results one year, one month, one week and one day before the election.

survey and the local trend specification. For the one week and the one day forecast all predictions are essentially coequal. It appears as if there is not a lot to gain in terms of predictive accuracy deploying DLM models to generate lower root mean square errors. But as our discussion of poll results reveal, the error is not the only front on which the choice of the prediction method should be decided. We also would like to have intervals that correctly include the final outcome.

The predictive intervals for election day are presented in Figure 4. The dots and intervals are shaded with the density/probability found around plus/minus one percentage point of the final outcome. Darker colors have heavier coverage of this area[15]. The one year a head forecasts exhibit the widest intervals, with small but existing coverage of

[15]In terms of the DLM predictive distributions, the densities can be interpreted as probabilities.

	2012-09-22	2013-08-22	2013-09-15	2013-09-21
survey	0.167	0.667	0.667	1.000
local level (w/o cov.)	1.000	0.833	1.000	1.000
local level (with cov.)	1.000	0.833	0.833	0.833
local trend (with cov.)	0.833	1.000	1.000	0.833

Table 2: Coverage of 2013 election result by 95% HPD forecasting intervals by model and forecasting date.

the final result. It is particular in this situation that the local level model appears well suited to form expectations about the final outcome. Both local lecel predictive intervals include all final election results. The coverage can be seen in Table 2. Although these intervals are large, they potentially represent what is possible within one year. For one month ahead, the forecasts get more accurate. Only for the local trend model the intervals a considerable wide. The surveys also get more predictive of the final outcome, generating stronger density around the final results compared to the forecasts from the local level model (0.348 versus 0.276). However, the confidence intervals fall short on including the final result (4 out of 6 are not included). Within the last weeks there is almost no difference between the model based forecasts and the surveys.

Overall, this first application shades some light in the predictive accuracy of DLM based forecasts. The model based estimates do not outperform the survey-based forecasts, but potentially generate a more accurate representation of predictive intervals for the final outcome. Years before the election expectable changes in support should result in relatively wide intervals that guarantee coverage of the outcome. While the model based forecasts are able to achieve this the confidence intervals are not. The application presents first evidence that beside generating potential interesting insights in the dynamics, extensions of the standard random walk model do not automatically enrich our ability to foretell the future. The findings have to be put in perspective, though. Here we only analyze four forecasts for one particular election campaign. A more detailed assessment of the predictive qualities should investigate a broader set of forecasts across different election campaigns.

5. Discussion

In this manuscript we discuss the application of DLM to the analysis of trial heat polls in multi-party elections. While most contributions in the literature rely on independent random-walk or local-trend specifications, we highlight that the DLM framework allows for further extensions that are valuable in modeling the evolution of party support during electoral campaigns. We further show how researchers can make use of a log-ratio transformation to deal with the constraint inherent in polling results. An application to the German Federal Elections 2013 reveals that polling the polls with DLMs offers a reliable method to form expectations about the election outcome. In particular, these models generate predictive intervals with large coverage even years before the election. However, the discussed extensions of DLMs do not necessarily increase our ability to predict the final result. They are potentiality more interesting to researchers who want to learn about specific dynamics of electoral support over the course of a campaign.

In a next step of this project, we plan to extend our forecast scenarios to include more than the four time-points. Weekly forecast within the last two years to the election would help to more accurately describe the evolution of uncertainty for the different models. Additionally, we hope to apply the discussed models to a larger variety of multi-party elections. Jennings and Wlezien [2013] compiled data of 312 electoral cycles which would allow us to asses the evolution variances and the predictive distribution in different contexts. We will thus be able to address questions about the validity of predictive intervals at different time points during a campaign.

We will also discuss a few potential extensions of the DLM specifications in the context of trial heat polls. In the specifications presented here, we ignored the fact that polls are administered by different polling institutes. House-effects are well-established and potentially make polling results systematically biased. Taking this into account should decrease the estimated evolution variance, which, in turn, could increase the precision of the predictive distributions. That being said, we also consider extensions that down-weight the impact of each data-point. The impact of new information of the model estimates in our application is generally very strong. From one perspective, this is sensible, as the margin-of-error for each poll is rather small. However, we are convinced that the translation from survey responses to electoral outcomes is more error-prone than the margin or errors suggests. An extended model would therefore allow for additional variance component that reflects the fact that polls are not elections.

References

John Aitchison. The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society. Series B. Methodological*, 44(2):139–177, 1982.

John Aitchison. *The statistical analysis of compositional data.* Chapman and Hall London, 1986.

Chris K Carter and Robert Kohn. On gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.

Cibele Queiroz Da-Silva and Guilherme Souza Rodrigues. Bayesian Dynamic Dirichlet Models. *Communications in Statistics - Simulation and Computation*, 44:787–818, 2015. ISSN 0361-0918. doi: 10.1080/03610918.2013.795592. URL `http://www.tandfonline.com/doi/abs/10.1080/03610918.2013.795592`.

Sylvia Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of time series analysis*, 15(2):183–202, 1994.

Simon Jackman. Pooling the polls over an election campaign. *Australian Journal of Political Science*, 40(4):499–517, 2005. ISSN 1036-1146. doi: 10.1080/10361140500302472.

Simon Jackman. The Predictive Power of Uniform Swing. *PS: Political Science & Politics*, 47(02):317–321, 2014. ISSN 1537-5935. doi: 10.1017/S1049096514000109.

Will Jennings and Christopher Wlezien. The Timeline of Election Campaigns: A Comparative Perspective. *Elections, Public Opinion and Parties*, 00(0):1–46, 2013. ISSN 00925853. doi: 10.1111/ajps.12189.

R E Kalman. A New Approach to Linear Filtering and Prediction Problems 1. *Journal of Fluids Engineering*, 82:35–45, 1960.

J N Katz and G King. A statistical model for multiparty electoral data. *American Political Science Review*, 93(1):15–32, 1999. ISSN 00030554. doi: 10.2307/2585758.

Drew a. Linzer. Dynamic Bayesian Forecasting of Presidential Elections in the States. *Journal of the American Statistical Association*, 108(501):124–134, 2013. ISSN 0162-1459. doi: 10.1080/01621459.2012.737735. URL `http://www.tandfonline.com.ezproxy.uct.ac.za/doi/full/10.1080/01621459.2012.737735{#}.U9evhkBkaeo`.

Kari Lock and Andrew Gelman. Bayesian combination of state polls and election forecasts. *Political Analysis*, 18(3):337–348, 2010. ISSN 10471987. doi: 10.1093/pan/mpq002.

Giovanni Petris. An R Package for Dynamic Linear Models. *Journal of Statistical Software*, 36(12):1–16, 2010. ISSN 15487660. doi: 10.1007/BF02887432. URL `http://www.jstatsoft.org/v36/i12/paper$\delimiter"026E30F$npapers2://publication/uuid/CBB963FA-1862-4D3C-BE12-8DD58B11DC28`.

Giovanni Petris, Sonia Petrone, and Patrizia Campagnoli. *Dynamic Linear Models with R*. 2008. ISBN 978-0-387-78170-9. doi: 10.1007/978-0-387-78171-6.

Andrew Q. Philips, Amanda Rutherford, and Guy D. Whitten. Dynamic Pie: A Strategy for Modeling Trade-Offs in Compositional Variables over Time. *American Journal of Political Science*, 60(1):268–283, 2016. ISSN 15405907. doi: 10.1111/ajps.12204.

S. E. Rigdon, S. H. Jacobson, W. K. Tam Cho, E. C. Sewell, and C. J. Rigdon. A Bayesian Prediction Model for the U.S. Presidential Election. *American Politics Research*, 37:700–724, 2009. ISSN 1532-673X. doi: 10.1177/1532673X08330670.

Neil Shephard. Partial non-gaussian state space. *Biometrika*, 81(1):115–131, 1994.

Emil Aas Stoltenberg. Bayesian Forecasting of Election Results in Multiparty Systems. (May), 2013. URL `https://www.duo.uio.no/handle/10852/36975`.

Michael Tomz, Joshua A. Tucker, and Jason Wittenberg. An Easy and Accurate Regression Model for Multiparty Electoral Data. *Political Analysis*, 10(1):66–83, 2002. ISSN 10471987. doi: 10.1093/pan/10.1.66.

Daniel Walther. Picking the winner(s): Forecasting elections in multiparty systems. *Electoral Studies*, 40:1–13, 2015. ISSN 02613794. doi: 10.1016/j.electstud.2015.06.003. URL `http://linkinghub.elsevier.com/retrieve/pii/S0261379415001109`.

Mike West and P Jeff Harrison. *Bayesian Forecasting and Dynamic Models, Bayesian Forecasting and Dynamic Models.* 1997.

Christopher Wlezien and Robert S Erikson. The Horse Race: What Polls Reveal as the Election Campaign Unfolds. *International Journal of Public Opinion Research*, 19(1), 2006.

A. Moments of additive logistic transformation

In order to break the compositional constraint of trial heat poll data, we use a transform – analyze – back transform strategy on polling data for each of K parties in a system[16]. For the untransformed data, the standard errors and covariances for the proportions y_k can be estimated as $\sigma_{y_k} = \sqrt{\frac{y_k(1-y_k)}{N}}$ and $Cov\,[y_k,\, y_q] = \sigma_{y_k,\, y_q} = \frac{-y_k y_q}{N}\ \forall\, k \neq q \in \{1, \ldots, K\}$. The complete variance-covariance matrix $\boldsymbol{V_Y}$ of the observed proportions is given by:

$$\boldsymbol{V_Y} = \begin{pmatrix} \sigma^2_{y_1} & \cdots & \sigma_{y_K,y_1} \\ \vdots & \ddots & \vdots \\ \sigma_{y_1,y_K} & \cdots & \sigma^2_{y_K} \end{pmatrix} \tag{16}$$

$$= \begin{pmatrix} \frac{y_1(1-y_1)}{N} & \cdots & \frac{-y_1 y_K}{N} \\ \vdots & \ddots & \vdots \\ \frac{-y_1 y_K}{N} & \cdots & \frac{y_K(1-y_K)}{N} \end{pmatrix} \tag{17}$$

This computation can naturally be deployed in the observational equation of a dynamic linear model. And in fact, many scholars rely on this trick to model changing support for political parties [Walther, 2015, Jackman, 2005].

For the log-ratio transformed shares $\boldsymbol{Y^*}$, the moments have to be adjusted. The variances and covariances of the non-lineary transformed data can be approximated using Taylor expansions. Using the first order Taylor expansion, we can approximate the expected value of $ln\left(\frac{y_k}{y_K}\right) = ln\,(y_k) - ln\,(y_K)$ at the means of y_k and y_K, μ_k and μ_K. We have

[16] Out of convince we drop the time subscript t here

$$\mathbb{E}\left[ln\left(\frac{y_k}{y_K}\right)\right] \approx \mathbb{E}\left[ln\left(\frac{\mu_k}{\mu_K}\right) + \mu_k^{-1}(y_k - \mu_k) - \mu_K^{-1}(y_K - \mu_K)\right] \tag{18}$$
$$= ln\left(\frac{\mu_k}{\mu_k}\right) + \mu_k^{-1}\mathbb{E}\left[(y_k - \mu_k)\right] - \mu_K^{-1}\mathbb{E}\left[(y_K - \mu_K)\right]$$
$$= ln\left(\frac{\mu_k}{\mu_K}\right)$$

To approximate the variance terms $Var\left[ln\left(\frac{y_k}{y_K}\right)\right]$, we make use of $Var\left[X\right] = \mathbb{E}\left[(X - \mathbb{E}\left[X\right])^2\right]$ and the first order approximation to $\mathbb{E}\left[ln\left(\frac{y_k}{y_K}\right)\right]$. The approximation to $Var\left[ln\left(\frac{y_k}{y_K}\right)\right]$ is:

$$Var\left[ln\left(\frac{y_k}{y_K}\right)\right] \approx \mathbb{E}\left[\left(ln\left(\frac{\mu_k}{\mu_K}\right) + \mu_k^{-1}(y_k - \mu_k) - \mu_K^{-1}(y_K - \mu_K) - ln\left(\frac{\mu_k}{\mu_K}\right)\right)^2\right] \tag{19}$$
$$= \mathbb{E}\left[\mu_k^{-2}(y_k - \mu_k)^2 - 2\mu_k^{-1}\mu_K^{-1}(y_k - \mu_k)(y_K - \mu_K) + \mu_K^{-2}(y_K - \mu_K)^2\right]$$
$$= \mu_k^{-2}\mathbb{E}\left[(y_k - \mu_k)^2\right] - 2\mu_k^{-1}\mu_K^{-1}\mathbb{E}\left[(y_k - \mu_k)(y_K - \mu_K)\right] + \mu_K^{-2}\mathbb{E}\left[(y_K - \mu_K)^2\right] \tag{20}$$
$$= \frac{\sigma_{y_k}^2}{\mu_k^2} + \frac{\sigma_{y_K}^2}{\mu_K^2} - \frac{2\sigma_{y_k,y_K}}{\mu_k\mu_K} \tag{21}$$

Given equation 17 this simplifies:

$$Var\left[ln\left(\frac{y_k}{y_K}\right)\right] \approx \frac{1}{N}\left[\frac{1 - y_k}{y_k} + \frac{1 - y_K}{y_K} + 2\right] \tag{22}$$

The covariance of X and Y is $Cov\left[X, Y\right] = \mathbb{E}\left[(X - \mathbb{E}\left[X\right])(Y - \mathbb{E}\left[Y\right])\right]$. Again, using the first order Taylor expansion to approximate the expected values of the transformed proportions, we obtain:

$$Cov\left[ln\left(\frac{y_k}{y_K}\right), ln\left(\frac{y_q}{y_K}\right)\right] \approx \mathbb{E}\left[\left(ln\left(\frac{y_k}{y_K}\right) - ln\left(\frac{\mu_k}{\mu_K}\right) - \mu_k^{-1}(y_k + \mu_k) - \mu_K^{-1}(y_K - \mu_K)\right)\ldots\right] \tag{23}$$

$$= -\frac{\sigma_{y_k,y_K}}{\mu_k\mu_K} - \frac{\sigma_{y_q,y_K}}{\mu_q\mu_K} + \frac{\sigma^2_{y_K}}{\mu_K^2} \tag{24}$$

From equation 17 this simplifies:

$$Cov\left[ln\left(\frac{y_k}{y_K}\right), ln\left(\frac{y_q}{y_K}\right)\right] \approx \frac{1}{N}\left[1 + \frac{1-y_K}{y_K}\right] \tag{25}$$

This allows to fill out the complete moments for the transformed data in equation 2.

B. Forward Filtering Backward Sampling algorithm in a Gibbs Sampler

This section describes a sampler to obtain draws from the marginal distributions of the parameters of interest for the Dynamic Linear Models specifications outlined in the main text $\{\boldsymbol{\theta_{0:T}}, \boldsymbol{W_\theta}\}$.It mostly follows Petris et al. [2008, section 4.4] strategy for simulated-based Bayesian inference in DLMs [see also Frühwirth-Schnatter, 1994, Carter and Kohn, 1994, Shephard, 1994]. The proposed algorithm relies on a forward-filtering and backward sampling procedure and an additional Gibbs step to take draws from the evolution variance. We implement the algorithm in R, building on the dlm package [Petris, 2010].

The center of the algorithm is a Forward Filtering Backward Sampling (FFBS) procedure, resulting in draws from the state vector for all time points $\{\boldsymbol{\theta_{0:T}}\}$. The filtering step of the FFBS is to run a Kalman Filter [Kalman, 1960] over the time-line, which requires defined observational variance ($\boldsymbol{V_t}$) as well as evolution variance ($\boldsymbol{W_\theta}$ in the main

Algorithm 1 FFBS: Forward Filtering Backward Sampling

1: Run Kalman Filter.
2: Draw $\boldsymbol{\theta_T} \sim \boldsymbol{N}(\boldsymbol{m_T}, \boldsymbol{C_T})$
3: **for** For $t = T-1, \ldots, 0$: **do**
4: Draw $\boldsymbol{\theta_t} \sim \boldsymbol{N}(\boldsymbol{m_t}, \boldsymbol{C_t})$
5: **end for**

text). The filtering step also requires to specifiy prior distribution for the initial states. Commonly, those are specied by a multivariate normal distribution $\boldsymbol{\theta_0} \sim \boldsymbol{N}(\boldsymbol{m_0}, \boldsymbol{C_0})$. Based on this, the following step updates the states for each time point according to the following equations:

$$\boldsymbol{m_t} = \boldsymbol{a_t} + \boldsymbol{R_t}\boldsymbol{F_t'}\boldsymbol{Q_t}^{-1}\boldsymbol{e_t} \tag{26}$$

$$\boldsymbol{C_t} = \boldsymbol{R_t} - \boldsymbol{R_t}\boldsymbol{F_t'}\boldsymbol{Q_t}^{-1}\boldsymbol{F_t}\boldsymbol{R_t} \tag{27}$$

where $\boldsymbol{a_t} = \boldsymbol{m_{t-1}}$, $\boldsymbol{R_t} = \boldsymbol{C_{t-1}} + \boldsymbol{W_\theta}$, $\boldsymbol{Q_t} = \boldsymbol{R_t} + \boldsymbol{V_t}$, and $\boldsymbol{e_t} = \boldsymbol{Y_t} - \boldsymbol{m_{t-1}}$. Given the final update at time point T, the algorithm samples values backwards $\boldsymbol{\theta_t} \sim \boldsymbol{N}(\boldsymbol{h_t}, \boldsymbol{H_t})$, starting at T, adjusting the distribution for every backward-step by:

$$\boldsymbol{h_t} = \boldsymbol{m_t} + \boldsymbol{C_t}\boldsymbol{G'_{t+1}}\boldsymbol{R^{-1}_{t+1}}(\boldsymbol{\theta_{t+1}} - \boldsymbol{m_{t+1}}) \tag{28}$$

$$\boldsymbol{H_t} = \boldsymbol{C_t} + \boldsymbol{C_t}\boldsymbol{G'_{t+1}}\boldsymbol{R^{-1}_{t+1}}\boldsymbol{G'_{t+1}}\boldsymbol{C_t} \tag{29}$$

This creates one vector of draws from the posteriori distribution of the states $\boldsymbol{\theta_{0:T}}$. Here, the updating equations are presented in their most-general form. Given the assumptions made in our specifications about the matrices $\{\boldsymbol{G_t}, \boldsymbol{F_t}, \boldsymbol{V_t}, \boldsymbol{W_t}\}$ those simplify accordingly. The FFBS algorithm is systematized in box 1.

The description highlights that obtaining draws for the states is conditional on the

Algorithm 2 FFBS in a Gibbs Sampler

1: Initialize: set $\boldsymbol{W_\theta} = \boldsymbol{W_{\theta^{(0)}}}$
2: **for** For $i = 1, \ldots, N$: **do**
3: Draw $\boldsymbol{\theta_{0:T}^{(i)}}$ from $P(\boldsymbol{\theta_{0:T}} \mid \boldsymbol{Y_{1:T}}, \boldsymbol{W_\theta} = \boldsymbol{W_\theta^{(i-1)}})$ using FFBS.
4: Draw $\boldsymbol{W_\theta^{(i)}}$ from $\boldsymbol{P(W_\theta \mid y_{1:T}}, \boldsymbol{\theta_{0:T}} = \boldsymbol{\theta_{0:T}^{(i)}})$
5: **end for**

evolution variance $\boldsymbol{W_\theta}$. However, this part of the model is generally unknown and of central interest to our application. Devising an additional Gibbs step that follows the FFBS algorithm permits to us to take draws from this conditional posteriori as well. This sequence of steps gives rise to a complete algorithm presented in box 2.

In this, the additional Gibbs step succeeds the draw from conditional distribution of the states using FFBS. The chain is initialed at a value for the the evolution variance $\boldsymbol{W_\theta} = \boldsymbol{W_\theta}^{(0)}$, followed by the algorithm described in box 2. The next step is to take a draw from the posteriori of $\boldsymbol{W_\theta}$. Given appropriate prior distributions the conditional densities can straightforwardly derived. As the approach slightly depends on the model specification of $\boldsymbol{W_\theta}$, we discuss the three model specifications of the main text in the following sequentially.

For the the *local-level model (without covariances)* we rely on independent Gamma distributions to express priors for all of the levels $\boldsymbol{\mu_t} = [\mu_{t1}, \ldots, \mu_{tK-1}]$. The model implies that the evolution variance $\boldsymbol{W_\mu} = diag(\boldsymbol{\sigma_\mu})$ are independent. Assigning Gamma priors $P(\sigma_k^\mu) \sim G(a_k^\mu, b_k^\mu)$, results in a conditional posteriori distributions that is then given by [Petris, 2010, see page 168]:

$$\sigma_k^\mu \sim G\left(a_k^\mu + \frac{T}{2}, b_k^\mu + \sum_{t=1}^{T}(\mu_{kt} + \mu_{kt-1})^2\right). \tag{30}$$

For the *local-trend model (without covariances)* we use the same independent Gamma formulation. The trends $\boldsymbol{\alpha_t} = [\alpha_{t1}, \ldots, \alpha_{tK-1}]$ with priors $P(\sigma_k^\alpha) \sim G(a_k^\alpha, b_k^\alpha)$ yield conditional posteriori distributions:

$$\sigma_k^\alpha \sim G\left(a_k^\alpha + \frac{T}{2}, b_k^\alpha + \sum_{t=1}^{T}(\alpha_{kt} + \alpha_{kt-1})^2\right). \tag{31}$$

For the *third model with covariance structures*, we rely on independent Wishart distributions to express our prior beliefs about $P(\boldsymbol{W_\mu})$ and $P(\boldsymbol{W_\alpha})$. We follow the same specification as in [Petris et al., 2008, p.173]. The Wishardt priors for the levels and trends are given by: $P(\boldsymbol{W_\mu}) \sim \mathcal{W}(\boldsymbol{(\delta_\mu + 1)/2}, \boldsymbol{V_\mu/2})$ and $P(\boldsymbol{W_\alpha}) \sim \mathcal{W}(\boldsymbol{(\delta_\alpha + 1)/2}, \boldsymbol{V_\alpha/2})$. While $\boldsymbol{V_\alpha}$ and $\boldsymbol{V_\mu}$ specify the prior beliefs about variance structure, the δ's also decide the weight of this information. Values close too 2 essentially imply a smaller weights on the prior specification. Based on this prior specification the full conditionals are given by:

$$\boldsymbol{W_\mu} \sim \mathcal{W}(\boldsymbol{(\delta_\mu + 1 + T)/2}, \boldsymbol{(V_\mu + SS_\mu)/2}), \tag{32}$$

where $\boldsymbol{SS_\mu} = \sum_{t=1}^{T}(\boldsymbol{\mu_t} - \boldsymbol{G\mu_{t-1}})(\boldsymbol{\mu_t} - \boldsymbol{G\mu_{t-1}})'$; and

$$\boldsymbol{W_\alpha} \sim \mathcal{W}(\boldsymbol{(\delta_\alpha + 1 + T)/2}, \boldsymbol{(V_\alpha + SS_\alpha)/2}), \tag{33}$$

where $\boldsymbol{SS_\alpha} = \sum_{t=1}^{T}(\boldsymbol{\alpha_t} - \boldsymbol{G\alpha_{t-1}})(\boldsymbol{\alpha_t} - \boldsymbol{G\alpha_{t-1}})'$.

The two steps in 2 are repeated a large number of times to approximate the posterior density of the DLM specification. It is recommended to also work with a large number of initial burn-in iterations, to ensure that the chain reached it's steady state.

C. DLM Estimates

D. Estimates of evolution variance

	SPD/CDU	FDP/CDU	LINKE/CDU	GRÜNE/CDU	OTHER/CDU
SPD/CDU	0.000625	0.000000	0.000000	0.000000	0.000000
FDP/CDU	0.000000	0.002285	0.000000	0.000000	0.000000
LINKE/CDU	0.000000	0.000000	0.000536	0.000000	0.000000
GRÜNE/CDU	0.000000	0.000000	0.000000	0.000714	0.000000
OTHER/CDU	0.000000	0.000000	0.000000	0.000000	0.009916

Table 3: Mean estimates of $\boldsymbol{W}_{\boldsymbol{\mu}}$ for local-level model without covariances on $\boldsymbol{W}_{\boldsymbol{\mu}}$

	SPD/CDU	FDP/CDU	LINKE/CDU	GRÜNE/CDU	OTHER/CDU
SPD/CDU	0.001479	0.002054	-0.001050	-0.000653	-0.003764
FDP/CDU	0.002054	0.004757	-0.001633	-0.001420	-0.006263
LINKE/CDU	-0.001050	-0.001633	0.001148	0.000564	0.003079
GRÜNE/CDU	-0.000653	-0.001420	0.000564	0.000873	0.001964
OTHER/CDU	-0.003764	-0.006263	0.003079	0.001964	0.012427

Table 4: Mean estimates of $\boldsymbol{W}_{\boldsymbol{\mu}}$ for local-level model with covariances on $\boldsymbol{W}_{\boldsymbol{\mu}}$

	SPD/CDU	FDP/CDU	LINKE/CDU	GRÜNE/CDU	OTHER/CDU
SPD/CDU	0.001890	0.002296	-0.001354	-0.000655	-0.004453
FDP/CDU	0.002296	0.005512	-0.001817	-0.001567	-0.006796
LINKE/CDU	-0.001354	-0.001817	0.001432	0.000590	0.003542
GRÜNE/CDU	-0.000655	-0.001567	0.000590	0.000839	0.001942
OTHER/CDU	-0.004453	-0.006796	0.003542	0.001942	0.013175

Table 5: Mean estimates of $\boldsymbol{W}_{\boldsymbol{\mu}}$ for local-trend model with covariances on $\boldsymbol{W}_{\boldsymbol{\mu}}$ and $\boldsymbol{W}_{\boldsymbol{\alpha}}$

	SPD/CDU	FDP/CDU	LINKE/CDU	GRÜNE/CDU	OTHER/CDU
SPD/CDU	0.000013	0.000002	-0.000002	-0.000000	-0.000004
FDP/CDU	0.000002	0.000021	-0.000001	-0.000002	-0.000005
LINKE/CDU	-0.000002	-0.000001	0.000014	0.000001	0.000002
GRÜNE/CDU	-0.000000	-0.000002	0.000001	0.000013	0.000001
OTHER/CDU	-0.000004	-0.000005	0.000002	0.000001	0.000025

Table 6: Mean estimates of $\boldsymbol{W}_{\boldsymbol{\alpha}}$ for local-trend model with covariances on $\boldsymbol{W}_{\boldsymbol{\mu}}$ and $\boldsymbol{W}_{\boldsymbol{\alpha}}$

E. Filtered trends

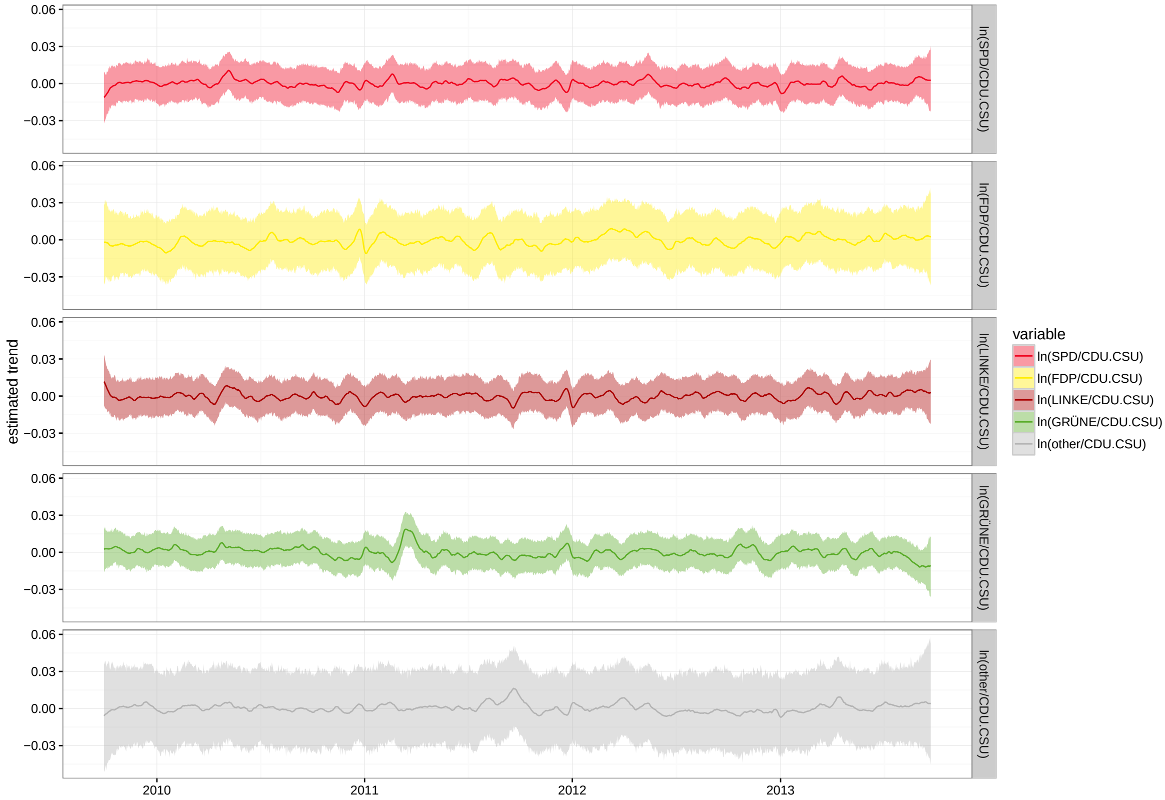

Figure 5: Evolution of trends calculated from