Beating the Odds - A State Space Model for predicting match results in the Australian Football League

Carien Leushuis*

Abstract

This thesis investigates whether state space models have the potential to predict the outcome of Australian Rules Football matches and can produce significant positive return over the bookmaker's odds. The point of departure is a sample of 18 Australian football teams over the period 2012 to 2016. Modeling and predicting a football match is a challenging task, since the model should incorporate two different random processes. Firstly, the evolution of parameters, i.e. the team strengths change over time as it incorporates changes in team composition, coaching, training ground, injuries etc. Secondly, the distribution of ranking data with these time-varying propensities changes stochastically over time. Given that we cannot observe all team-specific and location-specific factors, a dynamic state space model for team strengths is introduced. The team strengths are assumed to follow an order-one autoregressive process and are estimated using a recursive Kalman filter algorithm. Smoothed state estimates are applicable for ranking teams and predicting future outcomes of the matches. We show that beating the bookmaker's odds is a challenging task indicating that the betting markets are efficient.

Keywords: Sport Analytics; Kalman Filter Algorithm; State Space models; Predictive inference; Ranking; Australian Rules Football.

1 Introduction

1.1 The Secret to Sport success

Sport has an elevated place in Australia's national culture dating back to the early colonial period. For several years in a row, Melbourne has been crowned for being the sport capital of the world (Rolfe, 2016). The city Down Under is home to regular sell-out Australian Football matches, the Melbourne Cup horse racing, the Australian Open tennis Championships and each year Melbourne plays host to the Formula 1 Grand Prix. Melbourne prides itself on the calendar of world sports championships which has naturally

---

*Carien Leushuis received a bachelor degree in Econometrics & Operations Research at Maastricht University in 2017, where she currently takes the Master in the same field.
Contact: C.leushuis@student.maastrichtuniversity.nl

translated into the popularity of sport betting and gambling, with Australian Rules Football (AFL) leading the way. The collection of predictions is reflected by the bookmaker's odds, which tells how likely an event is to happen (Koopman & Lit, 2012).

The question is whether statistical models will be able to beat the odds. Ranking and predicting sport teams and their players is an emerging topic of statistical analysis that captures the interest of both professionals and fans. Is Hawthorn a better team than North Melbourne? What is the effect of home advantage on the outcome of an AFL game? Which teams represent good value from a betting perspective? Each of these issues relate to predicting outcomes in sports. Due to the highly stochastic nature, modeling and predicting sport results is a challenging task. In this paper, we are interested in the aforementioned questions. More specifically, this thesis investigates whether state space models have the potential to predict the outcome of Australian Rules Football matches and can produce significant positive return over the bookmaker's odds.

1.2 Trying to beat the Odds

Ranking sport teams based on their past performance has been extensively investigated in statistical literature dating back to the late 1970s. A ranking can be obtained by performing variations on the method of paired comparison models by Bradley and Terry (1952). Stefani (1977) considered the ranking of football teams as a simple linear regression model and proposed estimating the ratings by ordinary least squares. Harville (1977) extended this idea by constructing rankings for sport teams based on maximum likelihood estimation methods and Stefani (1980) showed that home-field advantage has a significant influence on the match results.

The models developed did not incorporate the stochastic nature of team performances. A notable innovation was the introduction of dynamic models to analyze team performances. Glickman and Stern (1998) developed the first dynamic state space model for National Football League game scores. In this study, Glickman and Stern (1998) assumed the team strength parameters to follow a first-order autoregressive process. Their approach in estimating team strengths that change over time is based on Bayesian inference. Over the years, different underlying distributions for match outcomes have been proposed. Maher (1982) used the double-Poisson distribution as underlying distribution for match results with the means expressing the team-varying goal scoring intensities of the two competing teams. Dixon and Coles (1997) extended this idea by including a dependence parameter in the double-Poisson model to capture correlations between home and away match scores. Owen (2011) used Markov chain Monte Carlo methods to study the time varying properties of football teams in continuous time. He found that the evolution of parameters for team performances and the effect of home-team advantage are more effectively analyzed using discrete time models. Koopman & Lit (2012) showed that football match results can be modeled using a bivariate Poisson distribution for pairs of match outcomes. They obtain maximum likelihood estimates for attack and defense strengths based on Monte Carlo simulation techniques.

Due to the nature of Australian football - the number of goals scored per match is relatively large - a linear model for the differences in scores can be used. This in contrast to regular Football, where the number of goals scored is relatively small and a Poisson (Skellam) distribution is necessary when modeling the (difference in) individual scores. The basis of our modeling approach was proposed by Manner (2015). In this study, basketball results are analysed based on a dynamic state space model for team strengths.

The observed differences in scores of the home team i and away team j are assumed to follow a Gaussian distribution. The statistical analysis is based on maximum likelihood estimation of parameters using the Kalman Filter algorithm. This paper applies a similar idea to modeling match outcomes in the Australian Football League.

1.3 Overview

The remainder of the paper is organized as follows. General knowledge about state space models and the kalman filter is provided in section 2. In section 3, we illustrate the methodology discussed on a high dimensional dataset consisting of Australian football match results for the seasons 2012 to 2016. In-sample and out-of-sample results of the dynamic model are discussed in section 4. The paper ends with a conclusion and recommendations for future research. Detailed empirical results for the smoothed teams strengths can be found in the appendix.

2 Statistical Modeling Framework

One of the most pervasive problems in statistics is inferring the driving forces behind economic variables which are unobservable and unmeasurable. Degrees of risk aversion, beliefs, abilities and other unmeasurable variables are particularly important when determining the decisions of economic agents. On an aggregate level, macroeconomic variables such as economic growth are driven by unobservable factors such as human capital accumulation and technological change. When explanatory variables are unobservable or unmeasurable, classical VAR models are no longer applicable when investigating the evolution of endogenous variables. State space models offer opportunities for allowing unobserved state variables to be incorporated into the observed model, containing VAR models as a special class.

2.1 State Space Models

State space models or dynamic models offer a framework for modeling several types of time series where the parameter can change over time. The idea was first introduced by Kalman (1960) and Kalman and Bucy (1961). They define the state space model to be characterized by two different random processes. First, there is the latent or state process x_t. The state process is assumed to be a Markov process, which implies that the probability distribution of the future states of the process conditional on both the present and past states depend solely upon the present state, i.e. $x_t|x_0, ..., x_{t-1} = x_t|x_{t-1}$. Therefore, the state prediction is a function of the prior state and the state error estimation. The essential assumption of the Kalman filter is that it is a linear system of equations and the error term follows a Gaussian distribution. Therefore, the basic linear Gaussian state space model with continuous states and discrete time intervals $t = 0, ..., T$ can be described as an order-one autoregressive state equation,

$$x_t = A_t x_{t-1} + w_t$$

where the disturbances $w_t \sim N(0, W_t)$ are serially independent and independent of the initial state vector. A_t is the evolution matrix at time t. This evolution matrix is often

block-diagonal where each block represents a certain aspect of the model, i.e. trends or seasonal effects. The Gaussian state space model is recursively defined starting from the initial prior distribution. This prior initial state vector is normally distributed with mean μ_0 and variance matrix Σ_0, that is $x_0 \sim N(\mu_0, \Sigma_0)$. We do not observe the state vector x_t directly, but only a linear transformed equation including noise which is considered to be the observation equation,

$$y_t = C_t x_t + v_t$$

where the measurement error is $v_t \sim N(0, V_t)$. The equation describes the relation between the observed time series y_t and (unobserved) latent variable x_t. C_t is the design matrix linking the state vector to the observation vector. Hence, dependence among observations is generated by the states. The state describes all the parameters we will have to predict, whereas the model describes how we think the system behaves as a linear function of the state (Douc et al., 2013). The premise of a state space model is that we have a (set of) states evolving over time. However, as the set of observations contain statistical noise, the 'true' states are not directly observable. As a consequence, the primary objective of state space modeling is to infer information about the states x as new information on observations y arrives. The system matrices A and C and the variances W and V are unknown and have to be estimated. Since the explanatory variables are unobservable, the usual least squares estimation is not appropriate.

2.2 Estimation of State Space models

A famous algorithm implementing this principle is the Kalman Filter. In general, the Kalman filter algorithm involves two stages, prediction and measurement update. In the prediction step the state at time t is predicted conditional on the past observations $y_{0:t-1}$. In the updating step the prediction is updated in light of a new observation y_t (Faragher, 2012). This section shows how the Kalman Filter and conjugate priors can be used to find an optimal solution to the state space model.

2.2.1 Initialization

The Kalman filter is a recursive algorithm for producing optimal linear forecasts for x_{t+1} and y_{t+1} based on the past information $Y_t = (y_0, ..., y_t)$. Let us assume that A, C, W and V are known. In a Kalman filter, the first measurement x_0 comes from sensor data where we define the prior mean,

$$\mu_0 = E(x_0) = \bar{x_0}$$

and prior covariance matrix,

$$\Sigma_0 = Var(x_0) = E(x_0 - \bar{x_0})(x_0 - \bar{x_0})'$$

2.2.2 Prediction

Since the state equation is assumed to be a Markov process, predictions for the future states can be based solely on the present state. Therefore, the prediction phase uses the state estimate from the previous time-step $\hat{x}_{t-1|t-1}$ only to obtain an estimate of

the current state. This estimate is known to be the a priori state estimate, since it solely incorporates information about the previously estimated state and information about observations is not yet incorporated. If the covariances are normally distributed, the predicted a priori state estimate conditional on the observed output minimizing the Mean-Square error is,

$$\hat{x}_{t|t-1} = E(x_t|y_1, ..., y_{t-1}) = A_t\hat{x}_{t-1|t-1}$$

and the corresponding predicted a priori covariance matrix,

$$\hat{\Sigma}_{t|t-1} = A_t\Sigma_{t-1|t-1}A_t' + W_t$$

These a priori predictions can be combined with the newly obtained observation to obtain an improved estimation of the state, which is called the a posteriori state estimate.

2.2.3 Measurement Update

In order to improve our understanding of the measurement update phase we rewrite the observation equation in terms of the state we want to estimate, $x_t = C_t^{-1}(y_t - v_t)$. The problem is that the noise v_t is unpredictable. Kalman (1960) came up with the idea to estimate the state by taking into account the current observation y_t and previously estimated state x_{t-1}. Therefore, when obtaining new observations, the predictions obtained above can be updated according to the Kalman (1960) formula, where the updated estimate of the current system state based on the current and past observations is given by,

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t(y_t - y_{t|t-1}) = \hat{x}_{t|t-1} + K_t(y_t - C_t\hat{x}_{t|t-1})$$

and posterior covariance matrix,

$$\Sigma_{t|t} = \Sigma_{t|t-1} - K_tC_t\Sigma_{t|t-1}$$

where

$$K_t = \Sigma_{t|t-1}C_t'(C_t\Sigma_{t|t-1}C_t' + V)^{-1}$$

which is defined as the Kalman gain. For simplicity let us disregard the observation matrix C. Note that if $K_t = 0$, then $\hat{x}_{t|t} = \hat{x}_{t|t-1}$ which implies that the newly obtained observation has no effect and the current state equals the previous state. If $K_t = 1$, then $\hat{x}_{t|t} = y_t$ which implies the previous state is irrelevant and we obtain the current state directly from the current observation. The actual Kalman gain will most likely be somewhere in between. The Kalman gain can be derived indirectly from the noise, since although we do not know the individual noise we know the average noise, V. Note that if the prediction error in the previous state $\Sigma_{t|t-1}$ was zero, the Kalman gain is zero and the next estimate equals the current state estimate since the prediction was accurate. If the prediction error equals 1, then the Kalman gain is $1/(1+V)$. If there is little noise in our observation equation, $V = 0$, the Kalman gain will be 1 and the state estimate x_k will be

strongly influenced by observation y_k. As the noise embedded in the observation equation grows larger, the noise prevents us from correcting bad predictions and the Kalman gain will become arbitrary small (Melenkevitz, 2016).

2.3 Forecasting with State Space models

When repeating the above prediction and updating steps we can obtain forecasts of the observations y_{t+T} by applying time updating on the states where $\hat{x}_{t+1|t} = A\hat{x}_{t|t}$ and $\Sigma_{t+1|t} = A\Sigma_{t|t}A^T + W$. Until this point we have assumed that the system matrices are known. Usually these hyperparameters have to be estimated by maximum likelihood estimation. The first step is to define the likelihood function to be maximized. Just like the Kalman filter calculates estimates of states recursively over time using incoming observations and a Bayesian updating process, we can calculate estimates of an unknown probability density function recursively over time. Define $y = (y_0, ..., y_T)'$, $x = (x_0, ..., x_T)'$ and let $\theta = \{A^*, C^*, W^*, V^*\}$ be the set of parameters to be estimated. Using Bayes rule we obtain the joint density $p(y, x|\theta) = p(y|x; \theta)p(x; \theta)$. Based on the Markov property it follows that the probability distribution over all states can be written as,

$$p(y|x; \theta) = \prod_{t=1}^{T} p(y_t|x_0, ..., x_t; \theta) = \prod_{t=1}^{T} p(y_t|x_t; \theta)$$

which implies the observation at the nth time-step is conditionally dependent on the current state (Koopman & Lit, 2012). In addition, by Bayes' rule we can factor the likelihood for the state process as,

$$p(x; \theta) = p(x_0; \theta)p(x_1|x_0; \theta)...p(x_t|x_0, ..., x_{t-1}; \theta) = p(x_0; \theta) \prod_{t=1}^{T} p(x_t|x_0, ..., x_{t-1}; \theta)$$

The likelihood associated with the joint density is the integral of the products of the probability distribution over all observations,

$$L(\theta|y) = \int p(y, x; \theta)dx = \int p(y|x; \theta)p(x; \theta)dx$$

The likelihood equation will in general be a complex function of the parameters such that maximization of the likelihood function with respect to θ has to be carried out using numerical maximization algorithms. If the likelihood functions is a continuous function, Gradient-based methods such as Newton's approximation algorithm can be applied, which is one of the most common iterative algorithms for solving least squares problems. In case of a large dataset, the Newton method may be inefficient due to the large number of iterations that have to be performed. Alternatively, Monte Carlo simulation methods can be used based on importance sampling to evaluate the likelihood function. Having obtained the maximum likelihood estimates for the hyperparameters, the state space model can be used to forecast the observations y_T for $T > t$.

3 Empirical Application

3.1 Data Description

In this section, the model discussed will be applied to a real world setting namely Australian Rules Football (AFL). The aim of the paper is to investigate whether state space models have the potential to predict the match results of AFL. The point of departure is a sample of 18 Australian football clubs over the period 2012 to 2016. This implies a total of five seasons running from March to August. Each team plays 22 matches per season resulting in a total of 198 matches per season. The AFL line up does not vary over the seasons, which makes modeling and ranking the data convenient. Most games have been played in the afternoons of Saturdays and Sundays, others are played on Friday evenings. The variation in number of goals scored is large, ranging from 22 to 193 goals per team per match. Therefore, a Gaussian dynamic state space model is used. In addition, a correction term for home team advantage is incorporated into the model. Since Australian rules football was invented in Melbourne, many teams like Geelong (1859), Carlton (1864), North Melbourne (1869), Essendon, Hawthorn and St Kilda (1873) are originally from different suburbs in Melbourne as shown in table 1. Shortly after the invention of the game, Australian rules Football began to spread to other Australian states like Tasmania, Queensland, New South Wales, South Australia and Western Australia. In the late 1980s national competitions followed under the name Australian Football League (AFL). The first South Australian team joined in 1991 and Fremantle was admitted in 1995. Nonetheless, Australian Rules football was (and remains) the dominant sport in the southern states, which includes Victoria, South Australia and Tasmania. A large number of games are still home to Melbourne Cricket Ground (MCG), which implies home advantage does not always apply as is the case in usual Football leagues. A more detailed explanation will be provided in the next sections.

3.2 AFL: The Rules of the Game

This section discusses the many facets of Australian Rules Football to provide an increased understanding on the game itself. AFL currently consists of 18 teams of which 10 teams are located in Victoria and two of each in South Australia, Western Australia, New South Wales and Queensland. Table 1 shows some background information in regard to the teams currently participating in AFL Season 2017, their training ground and number of winning premierships. AFL is played with an oval-shaped ball on a likewise oval-shaped field. A regular AFL games consists of two teams with 18 players, which are allowed to position anywhere on the field. The main objective is to obtain as many points by kicking the over-shaped ball between two tall goal posts. The ball can be moved forward in several ways by kicking, hand-balling, running and bouncing. A distinctive feature of AFL is that the player may run with the ball as long as the ball bounces or touches the ground at least once every 15 minutes. AFL is considered a contact sport, which implies players are allowed to tackle opposition players to obtain the ball. Failure to dispose the ball correctly results in a free kick for the opposition (Ryall, 2011). Each game consists of four 20 minute quarters and approximately 10 minutes of additional time per quarter. A goal is worth 6 points when the football is propelled through the goal posts by way of a kick. A so-called behind is worth 1 point, which means that the ball passes between a goal post and a behind post, if the ball hits the goal post or if a player sends the ball between the goal posts by touching it with any part of the body.

Jumper	Club	Location	Training Ground	Debut	Premierships
	Adelaide	Adelaide, South Australia	Adelaide Oval	1991	2
	Brisbane Lions	Brisbane, Queensland	The Gabba	1997	3
	Carlton	Melbourne, Victoria	MCG	1897	16
	Collingwood	Melbourne Victoria	MCG	1897	15
	Essendon	Melbourne, Victoria	Docklands Stadium	1897	16
	Fremantle	Perth, Western Australia	Subiaco Oval	1995	0
	Geelong	Melbourne, Victoria	Kardinia Park	1897	9
	Gold Coast	Gold Coast, Queensland	Carrara Stadium	2011	0
	Greater Western Sydney	New South Wales & Northern Territory	Sydney Showground	2012	0
	Hawthorn	Melbourne, Victoria	MCG	1925	13
	Melbourne	Melbourne, Victoria	MCG	1897	12
	North Melbourne	Melbourne, Victoria	Dockland Stadium	1925	4
	Port Adelaide	Adelaide, South Australia	Adelaide Oval	1997	1
	Richmond	Melbourne, Victoria	MCG	1908	10
	St Kilda	Melbourne, Victoria	Docklands Stadium	1897	1
	Sydney	Sydney, New South Wales	SCG	1897	5
	West Coast	Perth, Western Australia	Subiaco	1987	3
	Western Bulldogs	Melbourne, Victoria	Whitten Oval	1925	2

Table 1: Background and history of the 18 currently listed AFL clubs

3.3 Details of the Basic model

We consider a state space model in which the team performance is a time-varying latent state process. More specifically, it is assumed to follow a Gaussian autoregressive process of order one. Let T be the set of AFL teams in a particular season. Moreover, we define A_{ik} to be the number of goals scored by team $i \in T$ in game k and B_{jk} the number of goals scored by competing team $j \in T$ in game k, where $k = 1, ..., n$ is the index of the game and n is the total number of games in a particular season. Then we can define $y_{ijk} = A_{ik} - B_{jk}$ to be the difference in match outcomes of team i against team j in game k. As the total number of teams is denoted by T and each team plays a total of K games per season it follows that $n = \frac{T \times K}{2}$. In this particular model the number of teams is $T = 18$, the number of games played per team per season is $K = 22$ and the total number of matches per season is $n = 198$.

The difference in goals scored y_{ijk} is assumed to depend on the strengths or performances of team i and team j, denoted by β_i and β_j respectively. Moreover, when team i is playing at the home ground, a correction term δ is added to adjust for this advantage. In this paper we define home advantage to be the benefit that the home team is said to gain over the visiting team due to playing in their home city with their fans. The following model defines the outcome of the game,

$$y_{ijk} = \delta + \beta_{i,k} - \beta_{j,k} + v_{ijk}$$

where $v_{ijk} \sim N(0, \sigma_v^2)$. Home team advantage is considered to be the same for all teams and constant over time (Koopman & Lit, 2012). When both teams are playing at home, for instance Hawthorn is playing against North Melbourne at the Melbourne Cricket Ground (MCG), then δ equals zero. If team i is playing at the home ground while team j is playing on the away ground δ will be assigned a one.

The team strengths incorporate changes in team composition and will therefore in general not be constant over time. It is assumed that one team's composition is independent of how the other team composes their team. Moreover, team strengths may vary over time due to for instance different coaches, material, injuries and training ground. Therefore, for a particular team $i \in T$ strengths are assumed to follow and order-one autoregressive process,

$$\beta_{i,k} = \mu_i + \phi_i \beta_{i,k-1} + w_i$$

where μ_i is a constant term, ϕ_i is the autoregressive coefficient and the disturbances are independently and identically distributed with $w_i \sim N(0, \sigma_{w_i,}^2)$. The essential difference to the standard state space model as described in section 2 is the fact that the observations are not equidistant in calendar time and therefore the evolution of the team strengths is different from game to game (Manner, 2015). Nonetheless, the Kalman filter can be applied to estimate the model parameters and team strengths.

3.4 Implementation of the Kalman Filter

Before implementing the Kalman filter, the prior mean and variance have to be defined. The initial mean is set to $\beta_{i,1|0} = \mu_i/(1 - \phi_i)$ and initial covariance is $\Sigma_{i,1|0} = \sigma_{w_i}^2/(1 - \phi_i^2)$ for all individual teams, which are the conditional expectations of the predicted

team performances. Let $\beta_{i,k_i|k_i-1}$ be the predicted team strength of team i for game k_i conditional on the information obtained from game $k_i - 1$. Moreover, let $\beta_{i,k_i|k_i}$ denote the updated team strength conditional on the information obtained up to game k_i. The corresponding variances of β_{i,k_i} conditional on information at game $k_i - 1$ is denoted as $\Sigma_{i,k_i|k_i-1}$ and the updated variance is given by $\Sigma_{i,k_i|k_i}$. Based on Manner (2015), the Kalman filter procedure applied to the aforementioned dynamic state space model is defined as follows,

Prediction Step:

$$\beta_{i,k_i|k_i-1} = \mu_i + \phi_i \beta_{i,k_i-1|k_i-1}$$

$$\beta_{j,k_j|k_j-1} = \mu_j + \phi_j \beta_{j,k_j-1|k_j-1}$$

$$\Sigma_{i,k_i|k_i-1} = \phi_i^2 \Sigma_{i,k_i-1|k_i-1} + \sigma_{w_i}^2$$

$$\Sigma_{j,k_j|k_j-1} = \phi_j^2 \Sigma_{j,k_j-1|k_j-1} + \sigma_{w_j}^2$$

Observation Step:

$$\hat{y}_{ijk} = \delta + \beta_{i,k_i|k_i-1} - \beta_{j,k_j|k_j-1}$$

$$G_{ijk} = \Sigma_{i,k_i|k_i-1} + \Sigma_{j,k_j|k_j-1} + \sigma_v^2$$

$$\hat{e}_{ijk} = y_{ijk} - \hat{y}_{ijk}$$

Updating Step:

$$\beta_{i,k_i|k_i} = \beta_{i,k_i|k_i-1} + \hat{e}_{ijk} \Sigma_{i,k_i|k_i-1}/G_{ijk}$$

$$\beta_{j,k_j|k_j} = \beta_{j,k_j|k_j-1} + \hat{e}_{ijk} \Sigma_{j,k_j|k_j-1}/G_{ijk}$$

$$\Sigma_{i,k_i|k_i} = \Sigma_{i,k_i|k_i-1} - \Sigma_{i,k_i|k_i-1}^2/G_{ijk}$$

$$\Sigma_{j,k_j|k_j} = \Sigma_{j,k_j|k_j-1} - \Sigma_{j,k_j|k_j-1}^2/G_{ijk}$$

3.4.1 Estimation of initial hyperparameters

Before the Kalman filter can be applied, initial values for the set of hyperparameters consisting of μ_i, ϕ_i, σ_v^2 and $\sigma_{w_i}^2$ have to be estimated for all teams. These estimates will be based on parameter estimates of the static model using Ordinary Least Squares. Given the high number of free parameters, we restrict the auto-regressive coefficient to be the same for all team, that is, $\phi_i = \phi$. This restriction is not strong since we expect the strengths for all teams to evolve slowly over time (Koopman & Lit, 2012). Moreover, we consider imposing the restriction that the disturbance variances $\sigma_{w_i}^2$ are the same for all teams, thereby assuming homoskedasticity. The hyperparameter vector is given by,

$$\theta = (\sigma_{w_i}^2, \sigma_v^2, \phi, \mu_i, \delta)$$

We define the model in matrix form were $\mathbf{y}$ is an $n \times 1$ vector of differences in goals scored for every game k, $\mathbf{e}$ is the $n \times 1$ vector of error terms, $\beta = [\; \delta \quad \beta_1 \quad ... \quad \beta_{18} \;]'$ is the vector of parameter coefficients and $\mathbf{X}$ is the $n \times (t+1)$ design matrix. A particular row of this matrix has a 1 as its first element in case team i has a home advantage, 1 in column $i + 1$ and -1 in column $j + 1$ which denotes that team i is playing against team j. The remaining elements are assigned a value of zero. The model can be written as a basic linear regression model,

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{v}$$

However, the matrix $\mathbf{X}'\mathbf{X}$ is singular, i.e. its determinant equals zero and is therefore not invertible. There are several possible methods to make a singular matrix non-singular. One approach is to remove either the home team advantage variable or one of the dummies to avoid perfect multicollinearity. Alternatively, a bit of noise may be added to the singular matrix making the determinant go from zero to approximately zero, which in turn makes it possible to solve the system. This paper uses the former approach. We impose the restriction $\beta_1 = 0$ which results in the strength of the first team becoming the reference category. The parameters can be estimated using OLS,

$$\hat{\beta}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

This provides OLS estimates for the home team advantage coefficient δ and the strength of the teams $\beta_1, ..., \beta_{18}$. An initial estimate for ϕ and μ_i for $i = 1, ..., 18$ can be obtained by defining an autoregressive process of order one,

$$\beta_i = \mu_i + \phi\beta_{i-1} + w_i$$

The μ_i's and ϕ will be used an initial estimates for the individual teams in the state space model. An estimate for covariance terms σ_v^2 and σ_w^2 can be obtained by computing the sum of squared residuals and divided by their residual degrees of freedom.

3.4.2 Formulation of the log-likelihood function

Maximum likelihood estimates for the set of hyperparameters are found using the Kalman Filter by maximizing the log-likelihood function using numerical methods. The param-

eters in θ are restricted to the range $\sigma_w^2 > 0$, $\sigma_v^2 > 0$ and $|\phi| < 0.99$. Based on Manner (2015), the log-likelihood contribution of the *kth* game is given by,

$$l(k) = \frac{1}{2}ln(2\pi) + \frac{1}{2}ln(G_{ijk}) + \frac{\hat{e}^2_{ijk}}{2G_{ijk}}$$

There are several optimization methods available to find the parameter values that maximize the likelihood function. Three of the most popular algorithms are the Nelder-Mead, BFGS (Broyden-Fletcher-Goldfarb-Shanno) and SANN (Simulated-annealing). The BFGS method is an iterative gradient-based method that searches for a stationary point of a continuously differentiable function. Since we are dealing with a nonlinear likelihood function due to the necessary case distinction for the home team advantage, the BFGS method is not applicable. A more general method is the Nelder-Mead simplex algorithm, which only requires function evaluations and does not require gradient evaluations. Consequently, it can be applied to nonlinear functions for which the derivatives may be unknown. A potential disadvantage may be an increase in time required to find the maximum. The Nelder-Mead method fails to converge, the reason most likely be the high number of free parameters that have to be estimated (McKinnon, 1996). Like the Nelder-Mead algorithm, simulated-annealing uses only function evaluations and can therefore be applied to non-differentiable functions. SANN is based on the Metropolis-Hastings algorithm and works in a large search space. The main disadvantage is that this simulation method is again relatively slow. Restricting the set of hyperparameters is necessary for the algorithm to search in the correct range and obtain reasonable estimations. The predicted and filtered strengths and variances are obtained by running the Kalman filter again using the set of optimal hyperparameters.

3.5 Implementation of the Kalman Smoother

The predicted team strengths derived from the Kalman filter are volatile which leads us to smooth the team strengths using the Kalman smoother. Smoothed state estimates, denoted by $\beta_{i,k_i|K}$ are obtained by iterating on the whole sample going from the last to the first game using the formula based on Manner (2012),

$$\beta_{i,k_i|K} = \beta_{i,k_i|k_i} + \phi\frac{\Sigma_{i,k_i|k_i}}{\Sigma_{i,k_i+1|k_i}}(\beta_{i,k_i+1|K} - \beta_{i,k_i+1|k_i})$$

The Kalman smoother provides smoothed estimates of the states using the strengths and variances of the predicted and filtered states as derived in the previous section using the Kalman filter. The final value of the predicted strengths has been used as initial value for the smoothed estimate $\beta_{i,k_i+1|K}$.

4 Empirical Results

In this section, results of the dynamic model will be discussed. We consider the modeling of regular season data regarding the seasons 2012 to 2016. In-sample results are discussed in section 4.1 whereas an out-of-sample evaluation is presented in section 4.2.

4.1 In-sample evaluation

Considering the optimal hyperparameters, point estimates for the autoregressive parameter ϕ are close to 0.99, providing evidence of persistent time variation in team strengths within a single season. This in contrast to Manner (2015), who observes persistent time-variation only when multi-season data is used. Intuitively the result seems reasonable as one would expect the strengths of the teams to change throughout the season due to injuries, trades, changes in coaching, etc. The effect of home team advantage varies per season and is estimated to be approximately 5 to 10 points per game. Moreover, it has been observed that filtered variances decrease significantly over the season and no clear pattern is visible. High variances are possible both for high and low-performing teams.

The next step in the analysis is to investigate the smoothed team strengths which have been estimated using the optimal hyperparameters as initial starting values in the Kalman Filter after which the Kalman smoother is used. The team strengths have been normalized to add up to zero making the static and dynamic model comparable as suggested by Manner (2015) using the formula $\tilde{\beta}_i = \hat{\beta}_i - \frac{1}{t}\sum_{j=1}^{t}\hat{\beta}_j$. The figure in appendix 1 shows estimates of the time-varying smoothed team strengths based on the dynamic state space model (solid line) and the static results based on OLS (dashed line). The change in team strengths at the beginning of the season is significant which may be due the fact that changes in coaching, training ground and team composition are more likely to take place at the beginning of a new season. The static and dynamic strengths are relatively close to each other in most cases, although there are some significant outliers. The following provides a more detailed analysis and explanations for the patterns observed in appendix 1.

Adelaide experienced a relatively steady performance over the seasons. In 2012 the newly appointed coach Brenton Sanderson started off a successful era at the club. The club surprised the Australians by showing match results that were above expectations. Adelaide and Hawthorn played against each other in the First Preliminary Final after which Hawthorn, the leading team in the AFL 2012 season, won the game by just five points. At the start of the 2014 season, the club moved home matches to the newly developed Adelaide Oval. The football park remained the training ground of the club. The 2015 season started successfully with a 77-point win over North Melbourne which is visible in appendix 1 by the significant low performance of North Melbourne. Moreover, a coach transition took place leading to further success of the team. Unfortunately, new coach Phil Walsh died midway the season resulting in instability of the team. This may explain the slight decline in team strength in the middle of this season. Assistant coach Scott Camporeale was appointed as interim coach for the remainder of the season after which Don Pick took the lead from 2016 onwards.

The Brisbane Lions struggled through one disappointing season after the other as the club was plagued with injuries leading to a record of matches lost. However, this devastating performance is far below the usual excellent results of the team, as Brisbane Lions won three Premierships in a row from 2001 onwards. Leading into the 2012 season, only two players from the triple-premiership remained. At the end of the 2016 season injuries hit the club again which may explain the drop in team strength as is evident from the figures as given in appendix 1. For Carlton the most notable developments are the pronounced increases in team strengths halfway the 2015 season. Informing the actual data, it appears that except for two matches halfway the season, Carlton has lost every other match. After the dramatic performance in 2015, former Hawthorn coach Brendon Bolton was appointed for the 2016 season taking the club higher on the AFL ladder again.

This may explain the high jump moving from the 2015 to 2016 season as becomes evident from appendix 1.

Collingwood and Essendon's team strengths are shown to be relatively volatile over the course of the seasons. Looking into the history of the Collingwood Football Club, it becomes evident that the club has not been performing according to expectations resulting in continuous rotations of players and coaches throughout the seasons. Essendon faced a flying start during the 2012 season after which only one more match was won during the second half of the season. This event is in line with the significant drop in team strength as shown in appendix 1. In 2013, the club moved to a new venue in the suburb of Melbourne Airport. This did not help Essendon to get back on track as their players were accused of using performance-enhancing drugs. In the midst of the doping scandal, coach Mark Thompson took the lead, bring the club back to the finals resulting in a relatively steady development of team strengths over the 2014 season. In 2015 the coach left and investigations by the AFL anti-doping tribunal led to the decision of suspending twelve players for two years leading to a dramatic decline in performance in 2015. Although the most experienced players were absent, Essendon impressed the public throughout the end of the 2016 season as is evident from the appendix.

Fremantle experienced one of the most dramatic falls from grace in history as they lost their first 10 games in the 2015 season. Injuries lead to several top players missing a large amount of matches. Greater Western Sydney joined the Australian Football League from 2012 onwards. The only victory of the team in their first AFL season was a 34-point win over Port Adelaide. In the second season the club performed even worse by losing every single game. Greater Western Sydney managed to become part of the competition from its third season onwards. The club won two of their first three games including their biggest rival, the Sydney Swans. Despite a few notable losses halfway the 2016 season, last season was Greater Western Sydney's most successful one leading to a fourth place on the AFL ladder. This pattern is visible in the appendix. Overall, Hawthorn is one of the most successful clubs winning a total of thirteen premierships. Nonetheless, in 2015 Hawthorn started the season in irregular form winning only half of the matches in the first eights rounds before ending the season with eight consecutive victories. The 2016 season began with retirements of two triple premiership players. Moreover, the club lost one of their top players due to injuries resulting in limited scoring potential for Hawthorn and lower team strengths as shown in appendix 1.

The smoothed team strengths of the Sydney Swans still show a rather erratic pattern. Both in 2015 and 2016 a number of victories followed by several heavy defeats halfway the season. West Coast started in turmoil after which a new coach was admitted for the 2014 season leading to a more steady performance from 2016 onwards. A devastating season for the Western bulldogs in 2012 due to trade failures and injuries resulted in a decision to participate in the local Victoria Football League, which proved to be a tremendous success. The success also payed off in the national Australian Football League which is evident from the steep increase in team strength over the 2013 season. Although a period of injuries followed, the Bulldogs slowly started to improve their performances.

4.2   Out-of-sample evaluation

In this section, we investigate the forecasting performance of the dynamic model as described above. We carry out a one-step-ahead forecasting analysis using predicted values based on the Kalman filter to forecast the outcome of the nine matches taking place in

the last round of each season. The hyperparameter vector is estimated using the previous rounds of the corresponding season. Finally, we compare the prediction results with the Bookmaker's odds. The odds ratio's have been derived from www.aussportsbetting.com and are based on an average of 8 to 15 betting odds. Table 2 presents the fraction of games in which the correct winner was predicted. Overall, we find that approximately 60 to 64 percent of the match outcomes can be predicted correctly by the model, which is slightly lower than the performance of the Bookmaker's odds.

When investigating the out-of-sample forecasting performance, a total of 7 out of 9 matches have been correctly predicted by the dynamic state space model regarding the last round of the 2012 season. On September 2nd, Richmond was playing against Port Adelaide resulting in a draw, therefore incorrectly predicted both by the dynamic model and the odds. Geelong defended Sydney with a 34-point win on September 1st. Whereas the Bookmaker's correctly forecasted this outcome, it was not predicted correctly by the dynamic model resulting in a slightly lower percentage correctly predicted.

	In-Sample		Out-of-Sample	
	Dynamic	Odds	Dynamic	Odds
2012	0.608	0.737	0.777	0.888
2013	0.600	0.727	0.777	0.777
2014	0.621	0.680	0.666	0.666
2015	0.640	0.667	0.666	0.555
2016	0.606	0.682	0.777	0.666

Table 2: Fraction of correctly predicted results based on regular seasons matches

Considering the 2013 and 2014 season, an equal number of matches was correctly predicted both by the Bookmakers and the dynamic model. Whereas the match in which Adelaide defended the West Coast was correctly predicted by the dynamic model, the St Kilda against Fremantle match was correctly predicted by the Bookmaker's. In 2015 and 2016, the dynamic model performed slightly better by forecasting an additional correct match outcome over the Bookmakers. In general, the Bookmaker's odds remains a benchmark for prediction that appears to be very difficult to beat.

5 Discussion and Conclusion

The objective of this thesis was to investigate whether state space models have the potential to predict the outcome of Australian Rules Football matches and can produce significant positive return over the bookmaker's odds. Modeling and predicting a football match is a challenging task, since the model should incorporate two different random processes. Firstly, the evolution of parameters, i.e. the team strengths change over time as it incorporates changes in team composition, coaching, training ground, injuries etc. Secondly, the distribution of ranking data with these time-varying propensities changes stochastically over time. Given that we cannot observe all team-specific and location-specific factors, a Gaussian dynamic state space models for team strengths is introduced

for the analysis and forecasting of Australian Rules Football match outcomes in which the team strength is a time-varying latent state process.

The team strengths are estimated using a recursive Kalman filter algorithm that produces least squares optimal estimates for the team performance. Initial values for the set of hyperparameters are based on parameter estimates of the static model using OLS. Our empirical analysis is based on a dataset of match results based on five seasons of the Australian Football League. Evidence for persistent time variation in team strengths is found within the individual seasons. This is in line with the belief that the strengths of the teams change throughout the season due to injuries, trades, changes in coaching, etc. The home team advantage coefficient ranges from approximately 5 to 10 points per game. The smoothed team strengths show significant jumps at the beginning of each seasons, which confirms that trades, retirements and changes in coaching are more likely to take place at the start of a new season. The static and dynamic strengths are relatively close to each other in most cases, although there are some significant outliers. The forecasting performance of the model was evaluated using predicted values based on the Kalman filter, in which the last round of every season is used as an out-of-sample evaluation for forecasting Australian Football match results. Findings confirm that it is difficult to beat the Bookmaker's odds, indicating that the betting markets are efficient.

Future improvements can be made in different directions. Firstly, given the notable change in team strengths at the start of each season we could allow for a faster adjustment of team strengths at the beginning of the season. As suggested by Manner (2015), a dummy variable can be included for the variance in the first game of each season to capture the increase in variance at the beginning of the season. Secondly, the prediction error can be reduced by including more information about the matches. Potential explanatory variables for match outcomes are the duration between the matches and the traveling distance of the visiting team. This may result in an improvement of the forecasting performance. Thirdly, our statistical analysis is based on a classical perspective. A Bayesian approach can be taken to obtain predictive densities and an expanding window scheme can be used to produce forecasts for the full out-of-sample period.

Appendix Figure 1: Smoothed Team strengths over time
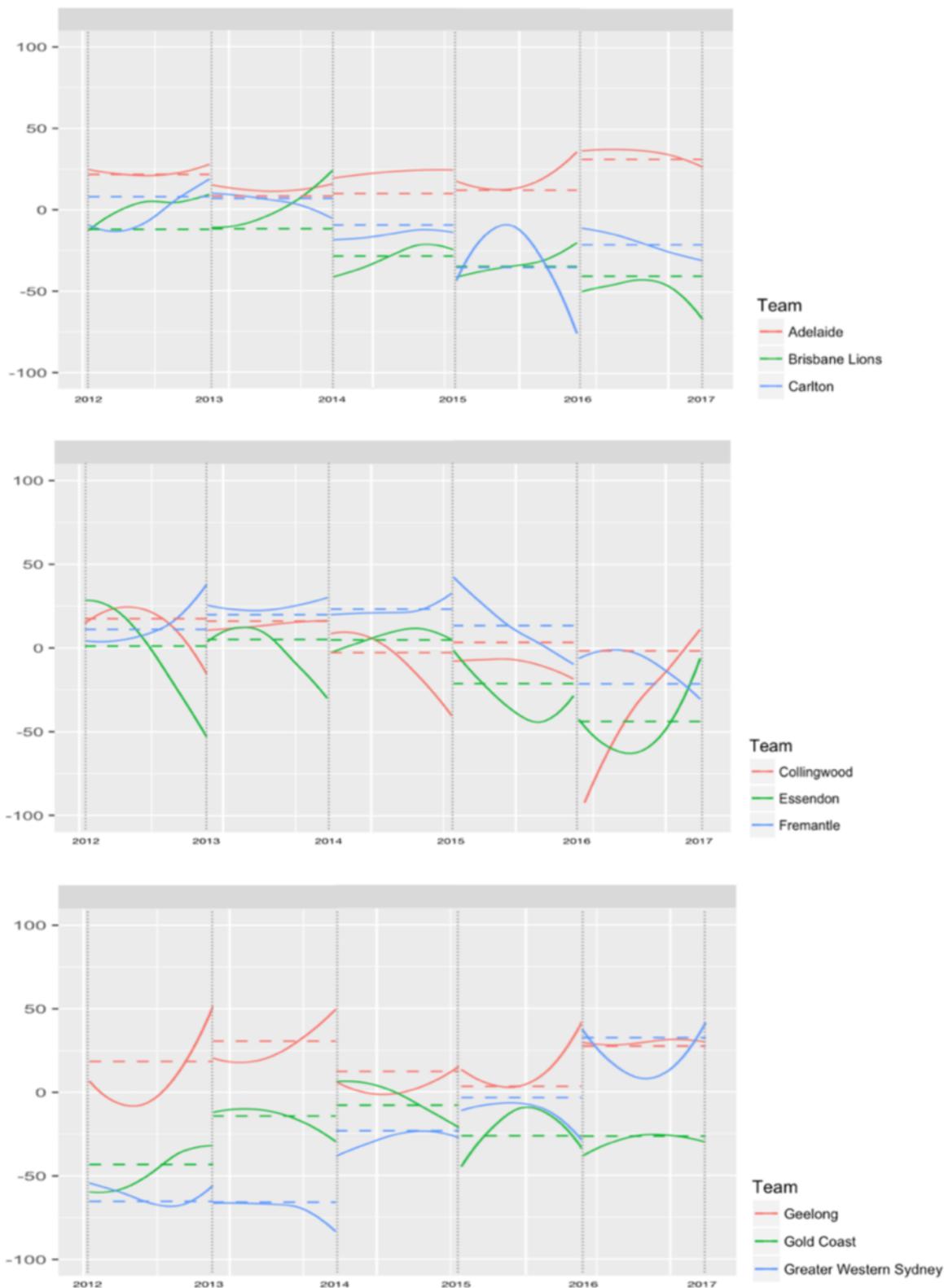
Note: Dynamic model: solid lines, static model: dashed lines

A State Space Model
for predicting match
results in AFL

17
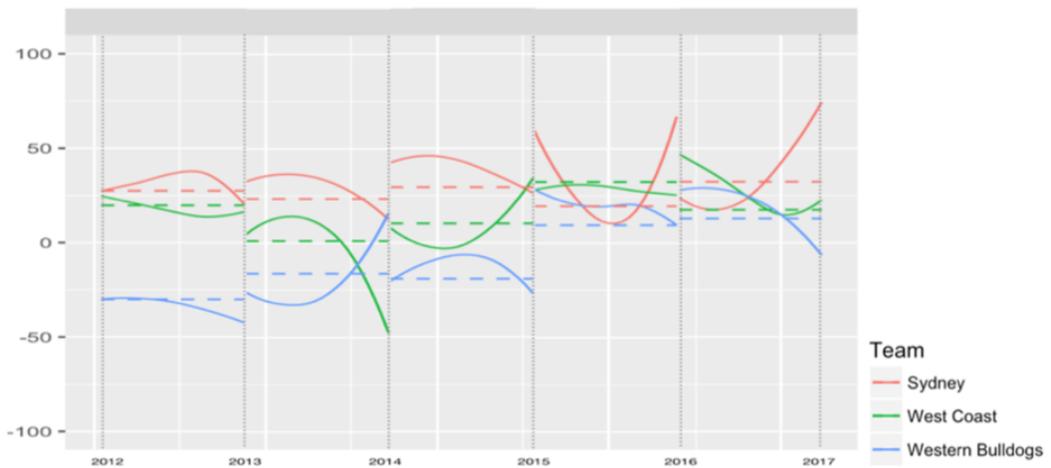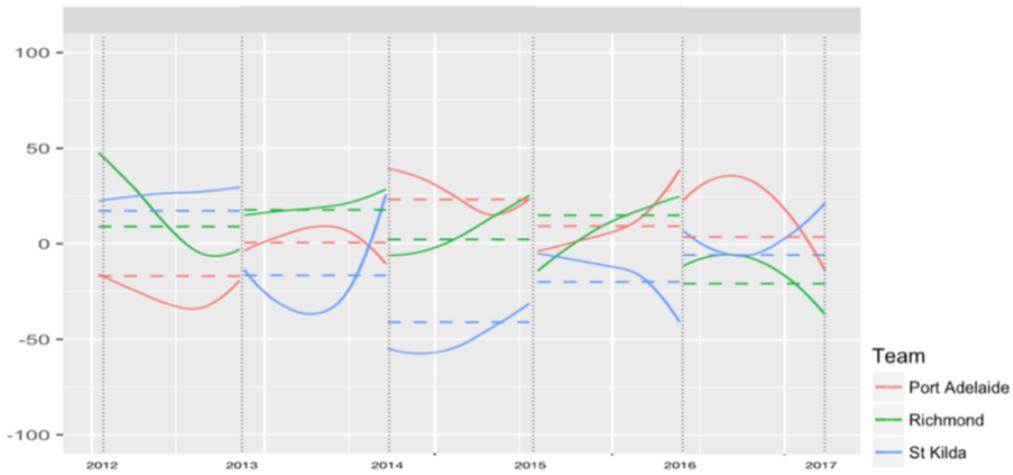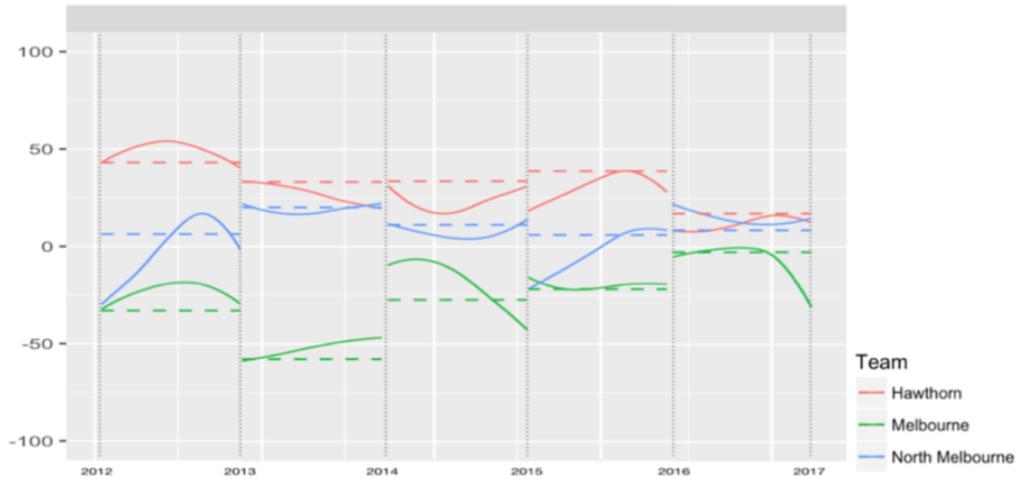
Note: Dynamic model: solid lines, static model: dashed lines

References

[1] Bradley, R.A. & Terry, M.E. (1952). Rank Analysis of Incomplete Block Designs: The Method of Paired Comparisons. *Biometrika*, 39(3/4), 324-345.

[2] Dixon, M.J. & Coles, S.G. (1997). Modeling association football scores and inefficiencies in the football betting market. *Applied Statistics*, 46(2), 265-280.

[3] Douc, R., Moulines, E. & Stoffer, D. (2013). *Nonlinear Time Series: Theory, Methods and Applications with R Examples.* London: CRC Press Taylor and Francis Group.

[4] Durbin, J. & Koopman, S.J. (2012). *Time series analysis by state space methods* (No. 38). Oxford University Press.

[5] Faragher, R. (2012). Understanding the Basis of the Kalman Filter Via a Simple and Intuitive Derivation. *Signal Processing Magazine, IEEE*, 29(5), 128-132.

[6] Glickman, M.E. & Stern H.S. (1998). A state-space model for national football league scores. *Journal of the American Statistical Association*, 93(441), 25-35.

[7] Harville, D. (1977). The use of Linear-Model Methodolgy to Rate High School or College Football Teams. *Journal of the American Statistical Association.* 72(358), 278-289.

[8] Kalman, R.E. (1960). A new approach to linear filtering and prediction problem. *Journal of Basic Engineering*, 82(1), 35-45.

[9] Kalman, R.E. & Bucy, R.S. (1961). New Results in Linear Filtering and Prediction Theory. *Journal of Basic Engineering*, 83(1), 95-108.

[10] Koopman, S.J. & Lit, R. (2012). A Dynamic Bivariate Poisson Model for Analysing and Forecasting Match Results in the English Premier League. *Journal of the Royal Statistical Society Series A*, 178(1).

[11] Maher, M.J. (1982). Modeling association football scores. *Statistica Neerlandica*, 36(3), 109-118.

[12] Manner, H. (2015). Modeling and forecasting the outcomes of NBA basketball games. *Institute of Econometrics and Statistics*, University of Cologne.

[13] McKinnon, K.I.M. (1996). Convergence of the Nelder Mead simplex method to a non stationary point.

[14] Melenkevitz, J. (2016). The Extended Kalman Filter: An Interactive Tutorial, Washington and Lee University.

[15] Owen, A. (2011). Dynamic bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter. *IMA Journal of Management Math-*

ematics, 22, 99-113.

[16] Rolfe, P. (2016, April 20). Melbourne crowned world's Sports City of Decade. *Sydney Herald Sun.*

[17] Ryall, R. (2011). Predicting Outcomes in Australian Rules Football (Doctoral dissertation). *RMIT University*, Melbourne, Australia.

[18] Stefani, R. T. (1977). Football and Basketball Predictions Using Least Squares. *IEEE Transactions on Systems, Man, and Cybernetics*, 10(2), 116-123.

[19] Stefani, R. T. (1980). Improved Least Squares Football Basketball, and Soccer Predictions, *IEEE Transactions on Systems, Man, and Cybernetics.* 10(2).