Dynamic Plackett-Luce Model for Ranked Data: Application to the Rating of Formula One Drivers*

Soichiro Yamauchi†
Harvard University
syamauchi@g.harvard.edu

This version: February 14, 2021
First draft: August 9, 2020

Abstract

Understanding the underlying popularity of items or ability of players from ranked data is a fundamental task in many fields such as computer science, economics, education research or sports analytics. This article proposes a dynamic model of ranked data for the rating estimation problem. Different from the existing methods, the proposed model can handle both the player-level and the team-level data, where multiple players can belong to the same team. Relying on the Polya-Gamma augmentation, I develop an efficient Gibbs sampler to estimate parameters of the model. The proposed model is applied to estimate the dynamic rating of Formula One drivers and racing teams from 1984 to 2019. I find that driver's career pattern differs even among top drivers and that disparities between teams have increased dramatically over the last decade.

Keywords: Formula One, Plackett-Luce model, ranked data, rating estimation

1 Introduction

The ranked-ordered data (or simply ranked data) arises whenever multiple items, players or students are sorted multiple times by sales, matches, or test scores. Understanding the latent features of such ranked items is a fundamental task in many fields such as political science (Gormley and Murphy, 2008), marketing (Bradlow and Fader, 2001), economics (Njenga, Onuonga and Sichei, 2018), or sports analytics (Murray, 2017). Specifically, the canonical Bradley-Terry model and its generalization (Plackett-Luce model) have been extensively used for estimating the rating in horse-racing (Plackett, 1975), car-racing (Hunter, 2004) and book-ranking (Caron and Teh, 2012), among others.

One of the common features in such ranked data is dynamics: Varying sets of players compete with each other over time. Motivated by a rating estimation problem of Formula One drivers and teams, where multiple teams and their players compete at a set of races each year, I develop a dynamic rating model of ranked data. The proposed model extends the standard Plackett-Luce (PL) model so that it allows for players to change their rating over time, and enables comparisons between players that do not directly

*I thank Ted Enamorado for valuable feedback. R package dyRank is available to implement the proposed method at https://github.com/sou412/dyRank.

†PhD candiate, Department of Government, Harvard University. URL: https://soichiroy.github.io/.

match up. I show that by log-transforming the original parameters in the PL model and introducing a set of indicator variables, the PL models can be represented as a multinomial logistic model with varying choice sets (e.g., Yamamoto, 2014).

The new multinomial formulation has several attractive features over the existing parametrization. First, we remove the positivity constraint of the model parameters. This is especially helpful when conducting a Bayesian inference, as constraints in the parameter space often lead to slower mixing. Second, extensions become simpler since we can rely on the vast literature on the multinomial logistic regression. For example, including match-level or item-level covariates to the rating model is now straightforward under the proposed parametrization. Finally, we can use the data augmentation strategy developed for the binomial likelihood to derive an efficient estimation algorithm (e.g., Polson, Scott and Windle, 2013; Linderman, Johnson and Adams, 2015). In Section 2.3, I develop a Gibbs sampler for the proposed dynamic rating model.

The proposed model is applied to the ranking data of drives and teams in Formula One racing since 1984. The naive application of the static model will pool the entire data and ignores the dynamical changes in driver's or team's rating. As any Formula One observer can attest, assuming the constant rating for racing teams (or drivers) is not appropriate as teams are under constant pressure from other competitors to improve their cars. The result shows that even among top drivers, the career path is quite different, thought there are several common patterns. Some drivers retire from the racing right after they reach their career peak (in terms of their rating), while some drives peak multiple times in their career. The analysis of the team ranking reveals that compared to a decade ago, teams are less competitive except for a few top teams, though there seems to be some improvements among mid-tier team in the 2019 season.

This paper is organized as follows. In the next section, I briefly review the canonical model for ranked data in a static setting. Section 2 describes the proposed dynamic model and its extension. Section 3 presents the empirical application of the proposed method to the Formula One data. Finally, I offer some concluding remarks in Section 4.

1.1 A Review of Models for Ranked Data

The original model for ranked data is developed for estimating ratings based on pair-wise competitions between players. Suppose that we observe a match between player i and player i'. The canonical Bradley-Terry model for the pair-wise competition is given by

$$\Pr(i \text{ beats } i') = \frac{\eta_i}{\eta_i + \eta_{i'}}$$

where the parameter $\eta_i > 0$ captures the latent "ability" or "popularity" of player i. Many algorithms have been proposed in the literature (see Hunter, 2004, and papers cited). Recently, Murray (2017) utilizes the re-parametrized version of the above model with $\lambda_i = \log(\eta_i)$, such that it can be expressed in a logistic form.

$$\Pr(i \text{ beats } i') = \frac{\exp(\lambda_i - \lambda_{i'})}{1 + \exp(\lambda_i - \lambda_{i'})}.$$

Murray (2017) then derives a Gibbs sampler based on the Polya-Gamma data augmentation (Polson, Scott and Windle, 2013).

The Plackett-Luce model generalizes the Bradley-Terry model to the multi-player setting where we observe a ranking of n players. Let $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_n)$ denote a ranking of players where $\rho_k \in \{1, \ldots, n\}$ indicates a player that ranks k in a match.

$$\Pr(\boldsymbol{\rho}) = \prod_{k=1}^{n} \frac{\eta_{\rho_k}}{\sum_{k'=k}^{n} \eta_{\rho_{k'}}}$$

The standard strategy is to place a Gamma prior on η_i and proceed with a Bayesian estimation. Caron and Doucet (2012) proposed a Gibbs sampler based on a novel latent variable formulation that exploits the properties of the exponential distribution. Virtanen and Girolami (2020) and Caron and Teh (2012) extend the formulation to the dynamic setting. Although the latent variable formulation based on the exponential distribution is elegant, it makes the extension of the model non-trivial. For example, Virtanen and Girolami (2020) uses a dynamic Gamma process to model the dynamic rating, but it is not clear how we can extend to model to other settings.

Instead, I show that the Plackett-Luce model can be casted as a Multinomial logistic model with varying choice sets. Although the formal treatment is given in the next section, we can preview the result in a simpler setting. Let λ_i denote a log transformation of η_i such that $\eta_i = \exp(\lambda_i)$. By the transformation, we eliminate the positivity constraint on η_i so that $\lambda_i \in \mathbb{R}$. By substituting $\exp(\lambda_i)$ for η_i, we have

$$\Pr(\boldsymbol{\rho}) = \prod_{k=1}^{n} \frac{\exp(\lambda_{\rho_k})}{\sum_{k'=k}^{n} \exp(\lambda_{\rho_{k'}})}$$

where again $\lambda_i \in \mathbb{R}$. Furthermore, the denominator term can be written as $\sum_{k'=k}^{n} \exp(\lambda_{\rho_{k'}}) = \sum_{i=1}^{n} \delta_{ik} \exp(\lambda_i)$ where δ_{ik} is an indicator variable that take 1 if unit i is ranked kth or below. We can see that this transformation turn the PL model into a product of n multinomial distributions.

2 The Proposed Methodology

2.1 Setup

We observe a ranking of $n_{mt} \leq N$ players at match $m \in \{1, \ldots, M\}$ in year $t \in \{1, \ldots, T\}$. For simplicity we assume that the number of matches are constant each year, so we observe in total MT ranked sets. Let $\boldsymbol{\rho}_{mt} = (\rho_{mt1}, \ldots, \rho_{mt,n_{mt}})^\top$ denote the ranking of players. In my application, ρ_{mtk} indicates a driver that finished kth in race m in year t.

We are interested in estimating the latent ability (or "rating") of N players. Let $\lambda_i^{(t)}$ denote the ith player's ability at time t. Here, I assume that player's rating is time-varying. Although sometimes it is possible that the player's ability itself does not vary much over time, the rating should reflect other factors that affect player's performance. For example, in many sports, changing a team often affects player's performance. Therefore, it is appropriate to assume that player's rating is time-varying. It is also clear that the time-varying formulation encompasses the time-constant model as a special case.

In many applications, players' career is shorter than the entire sample periods observed in the data. Since we are only interested in estimating the rating over the active career periods of player i, we define $\lambda_i^{(t)}$ for $t = \underline{T}_i, \ldots, \overline{T}_i$ where $\underline{T}_i$ and $\overline{T}_i$ denote the first and the last season, respectively.

2.2 Model

The dynamic Plackett-Luce models is defined on the observed ranking $\boldsymbol{\rho}_{mt}$.

$$\Pr(\boldsymbol{\rho}_{mt}) = \prod_{k=1}^{n_{mt}} \frac{\exp(\lambda^{(t)}_{\rho_{mtk}})}{\sum_{k'=k}^{n_{mt}} \exp(\lambda^{(t)}_{\rho_{mtk'}})} \tag{1}$$

where I use the re-parametrized verion of the PL model discussed in the previous section.

Proposition 1 (Multinomial Logit Representation of PLM)**.** Let $\delta_{imtk} = \mathbf{1}\{\sum_{k'=k}^{n_{mt}} \mathbf{1}\{\rho_{mtk'} = i\} = 1\}$ denote an indicator variable that takes 1 if player i is ranked kth or lower in matche m and time t. Then, the PL model in Equation (1) is reexpressed as

$$\Pr(\boldsymbol{\rho}_{mt}) = \prod_{k=1}^{n_{mt}} \prod_{i=1}^{N} \left\{ \frac{\exp(\lambda^{(t)}_{i})}{\sum_{i'=1}^{N} \delta_{i'mtk} \exp(\lambda^{(t)}_{i'})} \right\}^{\mathbf{1}\{\rho_{mtk}=i\}}.$$

Proof. The result follows from a simple algebra. The derivation is omitted. □

Proposition 1 says that the standard PL model after the re-parametrization can be written as the multinomial logit model with varying choice sets. The indicator variable δ_{imtk} controls the available set of choice items in match m, time t and rank k. This multinomial representation allows us to utilize wide machineries developed for the multinomial logit model and enables scholars to extend the PL model flexibly.

I propose a conditional sampling strategy that condition on other players' ability. It is certainly possible to derive a Gibbs sampler that samples the entire parameters at once, based the recent result on the Polya-Gamma augmentation for the multinomials (Linderman, Johnson and Adams, 2015), I choose the conditional approach as it is simpler to derive and implement in practice. for the multinomial distribution, which allows for estimating $\lambda^{(t)}_i$ jointly for all i, Conditional on other players' ability, we can write the above multinomial formula as the standard logistic form,

$$\begin{aligned}
\prod_{k=1}^{n_{mt}} \prod_{i=1}^{N} \left\{ \frac{\exp(\lambda^{(t)}_{i})}{\sum_{i'=1}^{N} \delta_{i'mtk} \exp(\lambda^{(t)}_{i'})} \right\}^{Y_{imtk}} &= \prod_{k=1}^{n_{mt}} \prod_{i=1}^{N} \left\{ \frac{\exp(\lambda^{(t)}_{i})}{\exp(\lambda^{(t)}_{i}) + \sum_{i'\neq i} \delta_{i'mtk} \exp(\lambda^{(t)}_{i'})} \right\}^{Y_{imtk}} \\
&= \prod_{k=1}^{n_{mt}} \prod_{i=1}^{N} \frac{\exp(\lambda^{(t)}_{i} - \log c_{imtk})^{Y_{imtk}}}{1 + \exp(\lambda^{(t)}_{i} - \log c_{imtk})}
\end{aligned}$$

where $c_{imtk} = \sum_{i'\neq i} \delta_{i'mtk} \exp(\lambda^{(t)}_{i'})$ and $Y_{imtk} = \mathbf{1}\{\rho_{mtk} = i\}$. Therefore, the contribution of unit i to the likelihood is given by

$$\begin{aligned}
&\prod_{m=1}^{M} \prod_{t=1}^{T} \prod_{k=1}^{n_{mt}} \left\{ \frac{\exp(\widetilde{\lambda}^{(t)}_{i})^{Y_{imtk}}}{1 + \exp(\widetilde{\lambda}^{(t)}_{i})} \right\}^{\delta_{imtk}} \\
&\quad \propto \prod_{m=1}^{M} \prod_{t=1}^{T} \prod_{k=1}^{n_{mt}} \left\{ \exp(\kappa_{imtk} \widetilde{\lambda}^{(t)}_{i}) \int_{0}^{\infty} \exp(-\omega_{imkt} (\widetilde{\lambda}^{(t)}_{i})^2/2) p(\omega_{imtk}) d\omega_{imtk} \right\}^{\delta_{imtk}}
\end{aligned}$$

where $\omega \sim \mathrm{PG}(1, 0)$ and the right-hand side is due to Theorem 1 of Polson, Scott and Windle (2013). Notice that the function inside of the exponential is quadratic in $\lambda^{(t)}_i$. Thus, the normal prior will be conjugate for λ parameter, which is different from the standard PL formulation that requires Gamma prior on the ability parameter.

Finally, we assume that the player's ability $\lambda_i^{(t)}$ follows the random-walk,

$$\lambda_i^{(t)} \sim \mathcal{N}(\lambda_i^{(t-1)}, \Delta_i^2)$$

and the initial condition is assumed to be $\lambda_i^{(0)} \sim \mathcal{N}(m_0, s_0^2)$. This formulation assume that the rating $\lambda_i^{(t)}$ moves smoothly conditional on one's ability in the previous period. The innovation term Δ_i^2 controls how variable $\lambda_i^{(t)}$ can be between time points.

2.3 Estimation

In this section, I develop a Gibbs sample for the dynamic Plackett-Luce model proposed in the previous section. I utilize the Polya-Gamma augmentation scheme for the logistic likelihood (Polson, Scott and Windle, 2013). As shown in the previous section, we arrive at the logistic likelihood from the multinomial representation by conditioning on other players' rating parameter $\boldsymbol{\lambda}_{-i} \equiv \{\boldsymbol{\lambda}_{i'}\}_{i'\neq i}$. The following Gibbs sampler exploits this structure and estimate one player's rating conditining on $\boldsymbol{\lambda}_{-i}$.

Data augmentation Compute $c_{imk} = \sum_{i'\neq i} \delta_{i'mtk} \exp(\lambda_{i'}^{(t)})$. Augment ω_{imtk} as follows

$$\omega_{imtk} \sim \text{Polya-Gamma}(1, \psi_{imtk})$$

where $\psi_{imkt} = \lambda_i^{(t)} - \log c_{imtk}$. Note that we need to augment ω_{imtk} only when $\delta_{imtk} = 1$.

Sampling the rating parameter I sample $\{\lambda_i^{(t)}\}_{t=\underline{T}_i}^{\overline{T}_i}$ via the forward-filtering and backward sampling (FFBS) algorithm (Frühwirth-Schnatter, 1994). To arrive at the FFBS algorithm, we first realize that the conditional posterior is given by the form

$$p(\{\lambda_i^{(t)}\}_{t=\underline{T}_i}^{\overline{T}_i} \mid \boldsymbol{\omega}_i, \boldsymbol{\rho}_i, \boldsymbol{\delta}_i, \boldsymbol{\lambda}_{-i}) \propto p(\{\lambda_i^{(t)}\}_{t=\underline{T}_i}^{\overline{T}_i}) \prod_{t=\underline{T}_i}^{\overline{T}_i} \exp\left\{ -\sum_{m=1}^{M}\sum_{k=1}^{n_{mt}} \delta_{imtk} \frac{\omega_{imtk}}{2}\Big(z_{imtk} - \lambda_i^{(t)}\Big)^2 \right\}$$

where $z_{imtk} = \kappa_{imtk}/\omega_{imtk} + \log c_{imtk}$ with $\kappa_{imtk} = \mathbf{1}\{\rho_{mtk} = i\} - 1/2$. Following the strategy proposed by Windle et al. (2013), we can rewrite the above as the dynamic linear model,

$$\mathbf{z}_{it} \sim \mathcal{N}(\lambda_i^{(t)}\mathbf{1}_t, \boldsymbol{\Omega}_{it}), \quad \lambda_i^{(t)} \sim \mathcal{N}(\lambda_i^{(t-1)}, \Delta_i^2)$$

where $\mathbf{z}_{it}$ is a vector of $\{z_{imtk}\}$ and $\boldsymbol{\Omega}_{it} = \text{diag}\{\omega_{imtk}\}$.

- **Forward filtering**: We compute the filtering density for $\lambda_i^{(t)}$, $p(\lambda^{(t)} \mid \mathbf{z}_{i,1:t})$, which is a Gaussian with mean $\mu_{t|t}$ and variance $\sigma_{t|t}^2$, $\lambda_i^{(t)} \sim \mathcal{N}(\mu_{t|t}, \sigma_{t|t}^2)$. The moments are computed as follows: for the mean, we have

$$\mu_{t|t} = \sigma_t^2(\mathbf{1}_t^\top \boldsymbol{\Omega}_{it}^{-1}\mathbf{z}_{it} + \mu_{t|t-1}/\sigma_{t|t-1}^2)$$

where $\mu_{t|t-1} = \mu_{t-1}$ and $\sigma^2_{t|t-1} = \sigma^2_{t-1} + \Delta^2_i$. The variance $\sigma^2_{t|t}$ is given by

$$\sigma^{-2}_{t|t} = \mathbf{1}_t^\top \boldsymbol{\Omega}_{it}^{-1}\mathbf{1}_t + \sigma^{-2}_{t|t-1}$$

The initial condition is given by $\lambda_i^{(0)} \sim \mathcal{N}(m_0, s_0^2)$.

- **Backward sampling**: We compute $p(\lambda_i^{(t)} \mid \lambda_i^{(t+1)}, \mathbf{z}_{i,1:t})$ for $t = T, \ldots, 1$ which is also a Gaussian with mean $\mu_{t|T}$ and variance $\sigma^2_{t|T}$. The updated moments are given by

$$\mu_{t|T} = (1 - a_t)\mu_{t|t} + a_{t+1}\lambda_i^{(t+1)}, \quad \text{and} \quad \sigma^2_{t|T} = (1 - a_t)\sigma^2_{t|t}$$

  where $a_t = \sigma^2_{t|t}/(\sigma^2_{t|t} + \Delta)$. The sequence is initialize by $\mu_{T|T}$ and $\sigma^2_{T|T}$.

  Given this, we sample $\lambda_i^{(t)}$ by $\lambda_i^{(t)} \sim \mathcal{N}(\mu_{t|T}, \sigma^2_{t|T})$.

For identification, I fixed one player's first-year rating $\lambda_1^{(1)} = 0$.

2.4 Extension: Estimating Team's Ability

In this section, I extend the proposed dynamic Plackett-Luce model for scaling teams. Let $G_i \in \{1, \ldots, J\}$ denote player i's team association. Then, the ranking outcome at the team level at match m and time t is given by $\boldsymbol{\pi}_{mt} = \mathbf{G}_{\boldsymbol{\rho}_{mt}}$. Here, $\pi_{mtk} \in [J]$ indicates a team that ranked kth at match m and time t. Note that this construction allows for the possibility that multiple players belong to the same team. For example, in my application of Formula One racing, each team has two racing drivers competing in a single race. This implies that when $G_{\rho_{mtk}} = G_{\rho_{mtk'}}$ for $k \neq k'$, we have that $\pi_{mtk} = \pi_{mtk'}$.

Let $\alpha_j^{(t)} \in \mathbb{R}$ denote the rating for team j at time t. To account for the possibility that a team might appear multiple times in a ranking $\boldsymbol{\pi}_{mt}$, I extend the model in the following form:

$$\Pr(\boldsymbol{\pi}_{mt}) = \prod_{k=1}^{n_{mt}} \frac{\exp(\alpha^{(t)}_{\pi_{mtk}})}{\sum_{j=1}^{J} \widetilde{\delta}_{jmtk} \exp(\alpha_j^{(t)})}$$

where $\widetilde{\delta}_{jmtk} = \mathbf{1}\big\{ \max\{\text{argmax}_{k'} \mathbf{1}\{\pi_{mtk'} = j\}\} \leq k\big\}$. The difference from the previous construction of the model is the introduction of a new indicator variable $\widetilde{\delta}_{jmtk}$. This variable properly accounts for the varying choice set, so that even if team j is selected, the team will not be immediately removed from the choice set as in the previous construction,

To gain an intuition how this indicator is constructed, consider a simple example where two teams with two players each competes in a match. Suppose that we observe $\boldsymbol{\pi} = \{1, 2, 2, 1\}$, which indicates that the players in the first team rank first and fourth, while the players in the second team rank second and third. In this example, we have $\widetilde{\delta}_{1,k} = 1$ for $k = 1, \ldots, 4$ because $\{1, 4\} = \text{argmin}_{k'} \mathbf{1}\{\pi_{k'} = 1\}$ and $4 = \max\{1, 4\}$. This example demonstrates that different from the model presented in the previous section, this model "allows" the first team to stay in the choice set, after being "picked" in the first choice.

After introducing the new indicator $\widetilde{\delta}$, the estimation is done identically as described in Section 2.3.

3 Empirical Application: Estimating the Rating of F1 Drivers

In this section, I apply the proposed dynamic Plackett-Luce model for the Formula One racing results to estimate drivers' (Section 3.2) and teams' rating (Section 3.3). I demonstrate that the proposed method can discover different career paths for top players, and enables comparisons between retired and current drivers or between past and current teams.

3.1 Background and Data Collection

Formula One is "the highest class of international single-seater auto racing" (Wikipedia). In each year, around 20 races take place internationally. In recent years, about 10 teams participate in the championship and each team have two main drivers, though the format varied in the past. Each race (called "Grand Prix") consists of the qualifying session, where drivers compete based on their lap times to determine the grid positions, and the main session, which is competed based on the fixed number of laps. This article focus on the ranking in the main session. Each driver and team is awarded points after each race reflecting their ranking in the race, and those who scores the most points at the end of the season wins the driver's and constructor's championship, respectively.

I collect all the Formula One Grand Prix (GP) results from a website (`f1-fansite.com`) through 1984 – a year that Ayrton Senna made his first appearance – to 2019. The data contains information about the final ranking of each driver and the driver's associated team at each GP. The data contains 14,694 unique observations where the average tenure of a team is 10.3 years and the average tenure of a driver is 5.8 years. There are a few teams that are observed throughout the sample periods such as Ferrari and McLaren, but no driver stayed in Formula One for the entire periods (the longest tenure is Michael Schumacher's 22 years). On average, a driver completes 48.6 races in their career.

There are several types of missingness in the outcome (i.e., ranking). The ranking is missing for a driver when a particular year is outside of the career period, that is, $t \notin [\underline{T}_i, \overline{T}_i]$. In addition, the ranking goes missing when a driver does not finish the race (DNF). This can happen for a variety of reasons such as a crash during the race or a retirement due to a mechanical failure. Obviously some of them are random, but it is also true that DNF reflects the driver's (or the team's) ability. Lastly, the driver (or the team) can be disqualified from a race. This can be due to a violation of technical regulations. In the following application, I treat them as "missing," ignoring differences in missing mechanisms for simplicity. A model that incorporates these features is left for the future work.

3.2 Estimating Drivers' Rating

I apply the proposed dynamic Plackett-Luce model to the driver ranking data. For the identification, I set the rating of Timo Glock's first year (2004) as zero. The initial condition of the rating is drawn from $\psi_i^{(0)} \sim \mathcal{N}(0, 0.5)$. I set the innovation term of the dynamic linear model as $\Delta_i^2 = \Delta^2 = 0.5$, which assumes that the rating is relatively smooth over time and does not dramatically varies. For the stability of the estimation, I dropped drivers whose results are missing more than 90% of their entries. This leaves me 190 unique drivers over 35 years of the Formula One history.

Figure 1 shows the estimated dynamic ratings for 14 world champions. Estimates for other drivers are presented in Appendix A. Estimates are obtained by running the Gibbs sampler for 4000 iterations in addition to the 1000 burin-in periods. I keep every 5th MCMC sample to reduce the auto-correlation between samples. Solid lines in the figure show the posterior median, while gray areas show the 95% credible intervals based on the posterior quantile. When a driver is absent from the Formula One racing during his career, I show the interpolated values during the absence (e.g., Schumacher was absent between 2007 and 2009).

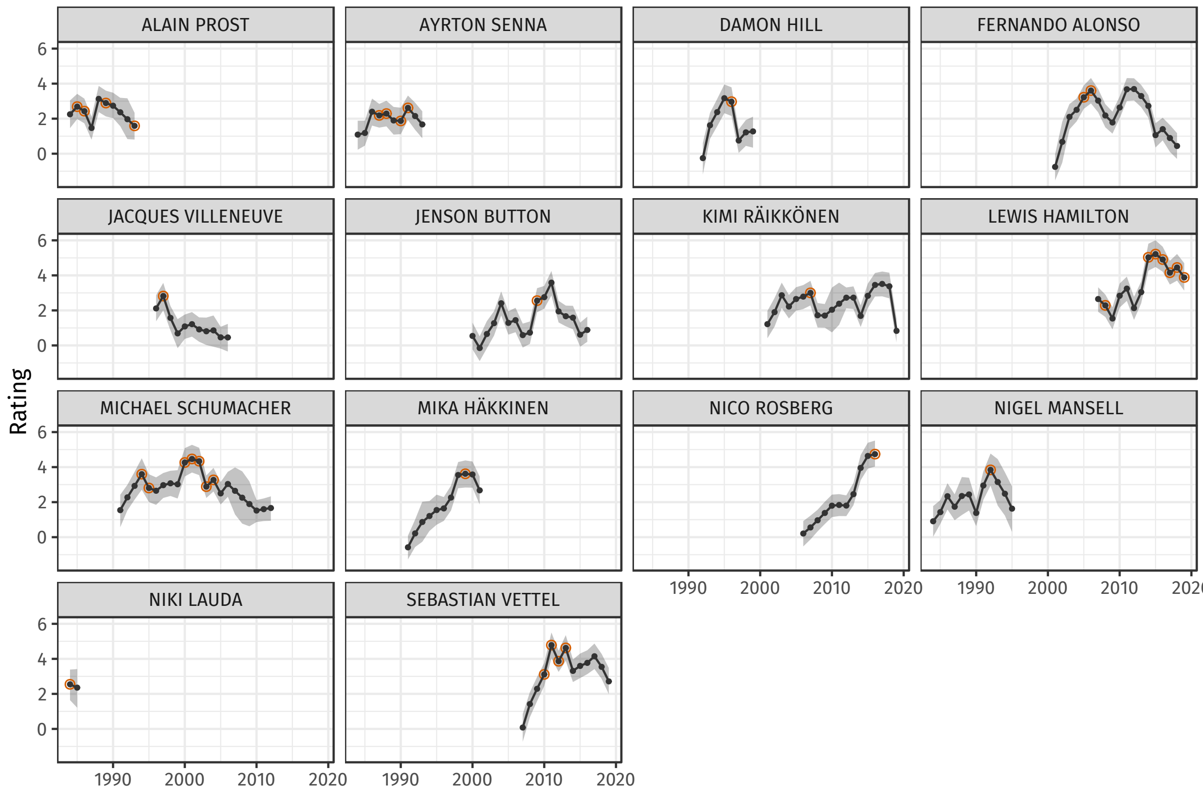


Figure 1: Estimated Ratings for World Champions (1984–2019). Solid lines are posterior medians and shaded areas are 95% credible intervals based on the posterior quantile. Orance circles indicate the year of the championship. The rating is relative to Timo Glock's first year (2004) which is fixed to zero for identification. Missing years are interpolated via predicted values (e.g., Schumacher: 2007–09, and Räikkönen: 2010–11).

Figure 1 shows a variation of career path across drivers. Some drives have a unimodal rating where they peak once in their career (e.g., Damon Hill or Nigel Mansell). Other drives such as Jenson Button and Fernando Alonso show a bimodal career path, while drives such as Nico Rosberg or Mika Häkkinen retire from the Formula One racing right after they reach the career peak.

Figure 2 shows the rating estimates of the top 25 drivers based on their highest rating in the career. Rating estimates for the world champions are shown in orange. We can see that world champions are clustered around the top drivers, though we see some drivers such as Mark Webber (2011) or Max Verstappen (2019) ranked higher than some of the world champions. I should note that these findings are not possible without a dynamic model where the comparison across time is possible. A static model or simple methods such as counting the number of wins do not admit across-year comparisons due to the differences in driver pools (i.e., who you are competing against is not fixed). The top ranked driver is

Lewis Hamilton (2015). In 2015 championship, Hamilton won 10 out of 19 GPs (11 pole positions) and became the world champion of the year.

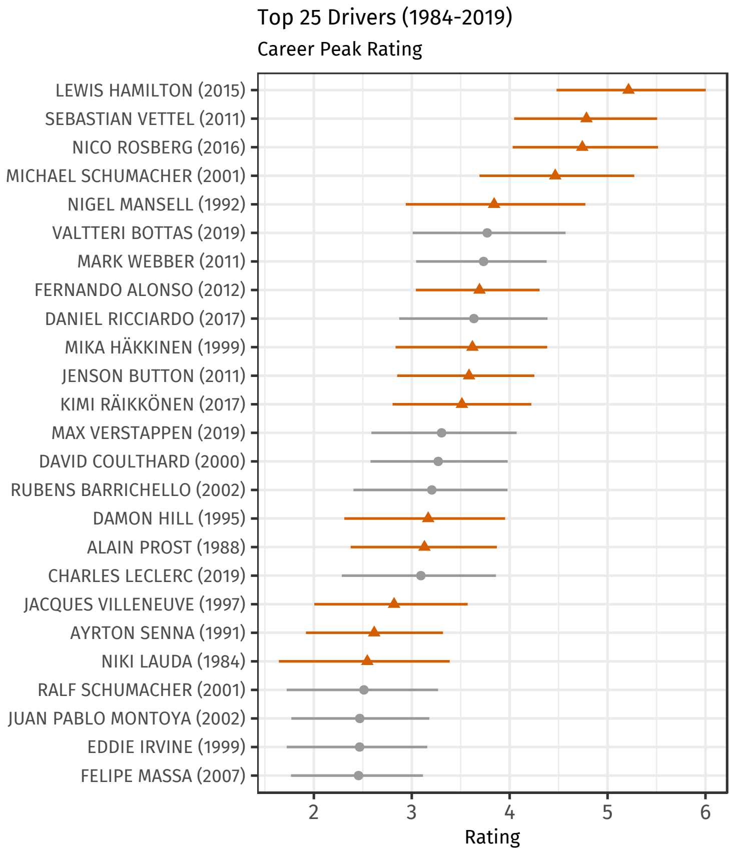


Figure 2: Estimated Rating with 95% Credible Intervals. Drivers are sorted based on the value of the posterior median. The estimates are based on the entire years, but the peak season is selected for presentation. Drivers in orange (triangle) are world champions (the peak year is not necessarily the championship year).

Finally, I emphasize that the rating estimates are not the pure reflection of driver's ability to drive. The performance in a race highly depends on the team's ability to develop a faster car and provide an accurate strategy. Thus, the rating should be interpreted as a score that incorporates many factors that affect the driver's standing in the championship. In fact, sudden shifts in a driver's rating are often associated with his move to another team. For example, in Figure 1, Kimi Räikkönen's rating plummeted in 2019. This possibly reflects his move from Ferrari to Alfa Romeo (Sauber) in 2019, where the latter team is not as competitive as Ferrari.

3.3 Estimating Teams' Rating

I apply the extension of the dynamic Plackett-Luce model discussed in Section 2.4 to the team ranking data. I dropped several teams that do not have sufficient number of race finishes. I also combine several teams that are considered to be a direct predecessor or successor. This creates a team that has a longer tenure in the Formula One racing, which helps improve the stability of estimation. For example, I combined Sauber, BMW Sauber and Alfa Romeo into a single team. To preserve the interpretability, however, I merge the current team with the previous team only when the number of observations is too few. This means that even though I combine BAR, Honda and Brawn into a single team, Mercedes is treated as a different team. Among the current teams, Racing Point is combined with Force India, in addition to Alfa Romeo's merge with Sauber. Haas is left as is because the predecessor is unclear. This operation leaves me 46 unique teams between 1984 and 2019.

Initial conditions are set as in the driver's rating estimation, and I set $\Delta_i^2 = \Delta^2 = 0.5$ for the innovation in the rating. The Gibbs sampler is run for 4000 iterations in addition to the 1000 burin iterations. I keep MCMC samples for every 5th iteration after the burin period.

Figure 3 shows the estimates of the dynamic rating of current constructors. Other estimates are available in Appendix B. The figure shows that some of the current teams show the downward trend since mid 2000's, while top tier teams such as Mercedes, Red Bull or Ferrari do not seem to exhibit such trends. This might imply that the current racing environment is less competitive than what it was 15 years ago as the disparities between teams increase over years. The highest point in the last decade was scored by Mercedes in 2015. As discussed in the drivers' rating section, this year saw Lewis Hamilton, one of the two Mercedes' drivers, winning the world championship. In fact, Mercedes dominated the entire GPs of the year; together with another team mate Nico Rosberg, Mercedes won 16 out of the 19 races.

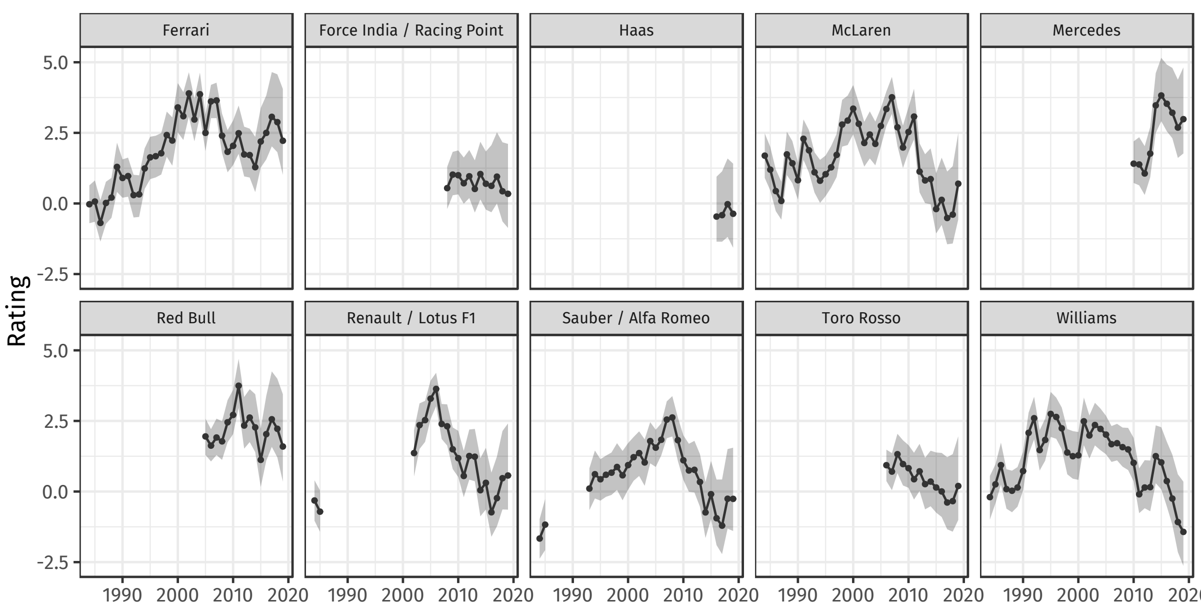


Figure 3: Dynamic Rating Estimates for the Current Formula One Teams. Estimates are based on the entire teams in the data (1984–2019), but current teams are presented in the figure. The rating is relative to the first year of Super Aguri (2006), which is fixed to 0. Solid lines show the posterior medians and gray areas show the quantile-based 95% credible intervals.

4 Concluding Remarks

The ranked data with time index is quite common in many applications. This paper has proposed a dynamic rating model to analyze such data. The proposed model allows the data to be at the player level or aggregated at the team level where multiple players can belong to the same team. An efficient Markov chain Monte Carlo algorithm is developed to estimate model parameters. I have applied the proposed model to the ranking data from Formula One racing.

References

Bradlow, Eric T and Peter S Fader. 2001. "A Bayesian lifetime model for the Hot 100 Billboard songs." *Journal of the American Statistical Association* 96(454):368–381.

Caron, Francois and Arnaud Doucet. 2012. "Efficient Bayesian inference for generalized Bradley–Terry models." *Journal of Computational and Graphical Statistics* 21(1):174–196.

Caron, François and Yee W Teh. 2012. Bayesian nonparametric models for ranked data. In *Advances in Neural Information Processing Systems*. pp. 1520–1528.

Frühwirth-Schnatter, Sylvia. 1994. "Data augmentation and dynamic linear models." *Journal of Time Series Analysis* 15(2):183–202.

Gormley, Isobel Claire and Thomas Brendan Murphy. 2008. "Exploring voting blocs within the Irish electorate: A mixture modeling approach." *Journal of the American Statistical Association* 103(483):1014–1027.

Hunter, David R. 2004. "MM algorithms for generalized Bradley-Terry models." *The Annals of Statistics* 32(1):384–406.

Linderman, Scott W., Matthew J. Johnson and Ryan P. Adams. 2015. Dependent Multinomial Models Made Easy: Stick Breaking with the PóLya-Gamma Augmentation. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*. NIPS15 Cambridge, MA, USA: MIT Press pp. 3456–3464.

Murray, Thomas A. 2017. "Ranking ultimate teams using a Bayesian score-augmented win-loss model." *Journal of Quantitative Analysis in Sports* 13(2):63–78.

Njenga, Githinji, Susan M Onuonga and Moses Muse Sichei. 2018. "Institutions effect on households savings in Kenya: A ranked ordered multinomial/conditional probit model approach." *Journal of Economics and International Finance* 10(5):43–57.

Plackett, Robin L. 1975. "The analysis of permutations." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 24(2):193–202.

Polson, Nicholas G, James G Scott and Jesse Windle. 2013. "Bayesian inference for logistic models using Pólya–Gamma latent variables." *Journal of the American Statistical Association* 108(504):1339–1349.

Virtanen, Seppo and Mark Girolami. 2020. Dynamic content based ranking. In *International Conference on Artificial Intelligence and Statistics*. pp. 2315–2324.

Windle, Jesse, Carlos M Carvalho, James G Scott and Liang Sun. 2013. “Efficient data augmentation in dynamic models for binary and count data.” *arXiv preprint arXiv:1308.0774* .

Yamamoto, Teppei. 2014. “A Multinomial Response Model for Varying Choice Sets, with Application to Partially Contested Multiparty Elections.” *http://web.mit.edu/teppei/www/research/dchoice.pdf* .

Appendix

A Additional Results: Drivers' Rating

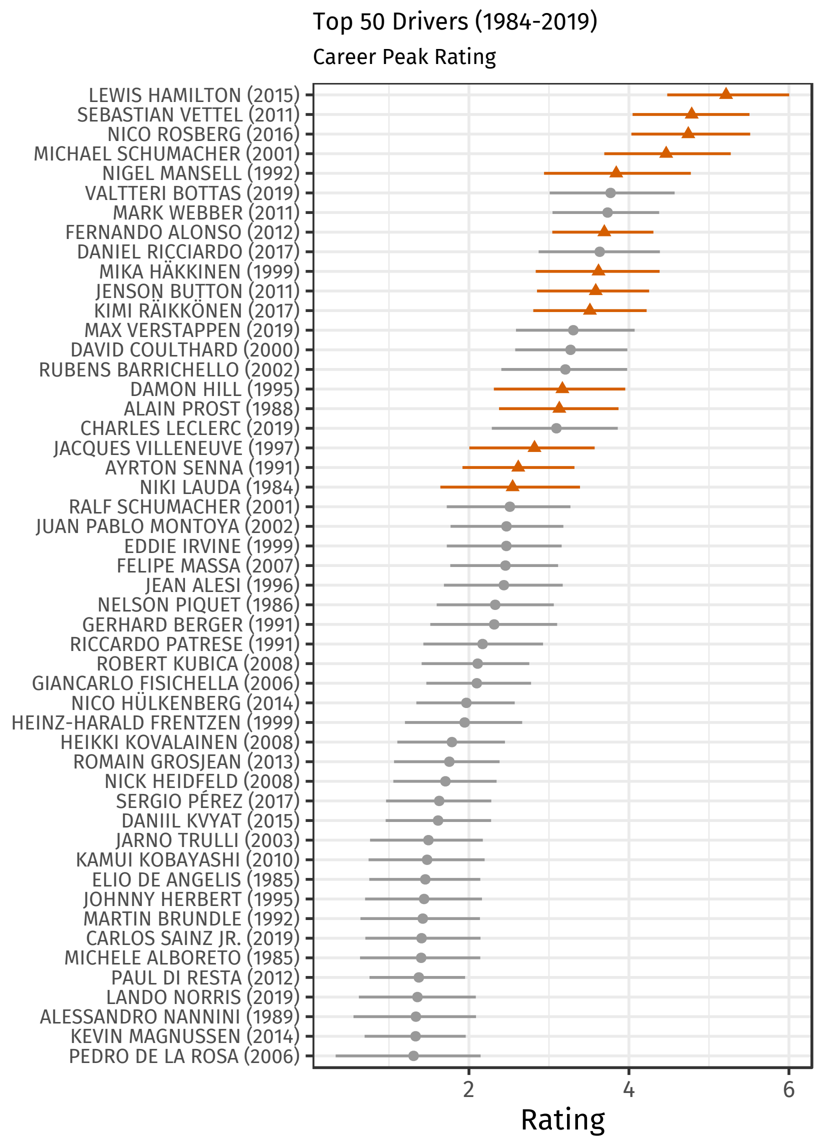
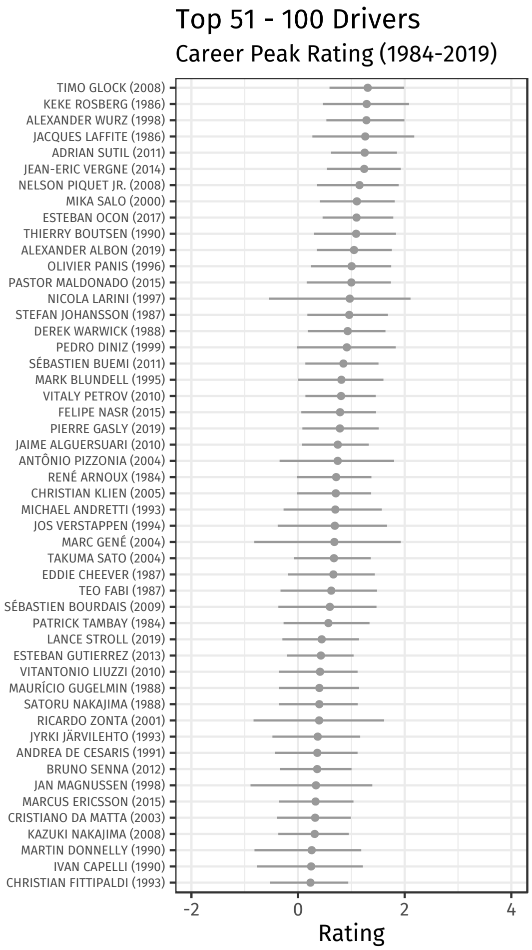


Figure A.1: Career peak ratings for top 50 drivers (1984–2019).



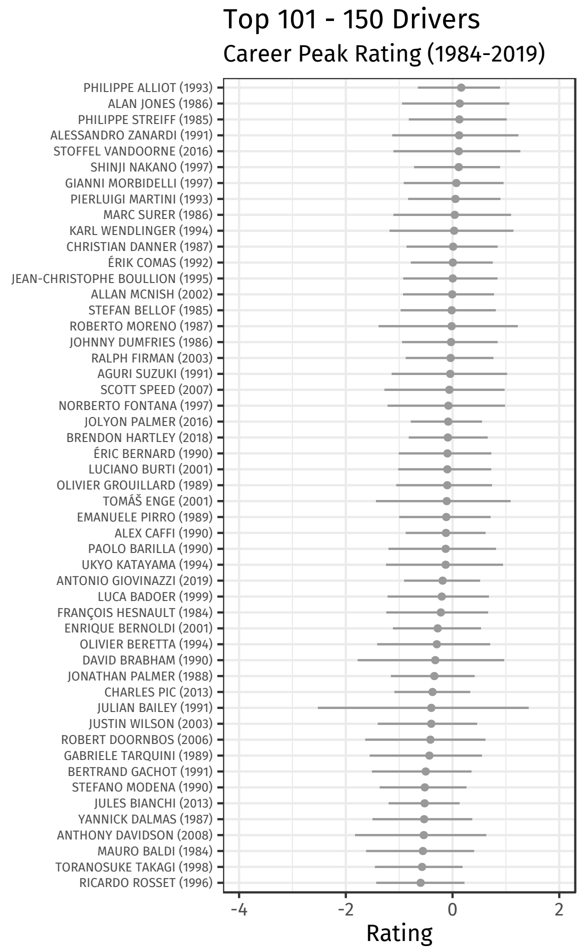


Figure A.2: Career peak ratings for top 51 to 150 drivers (1984–2019).

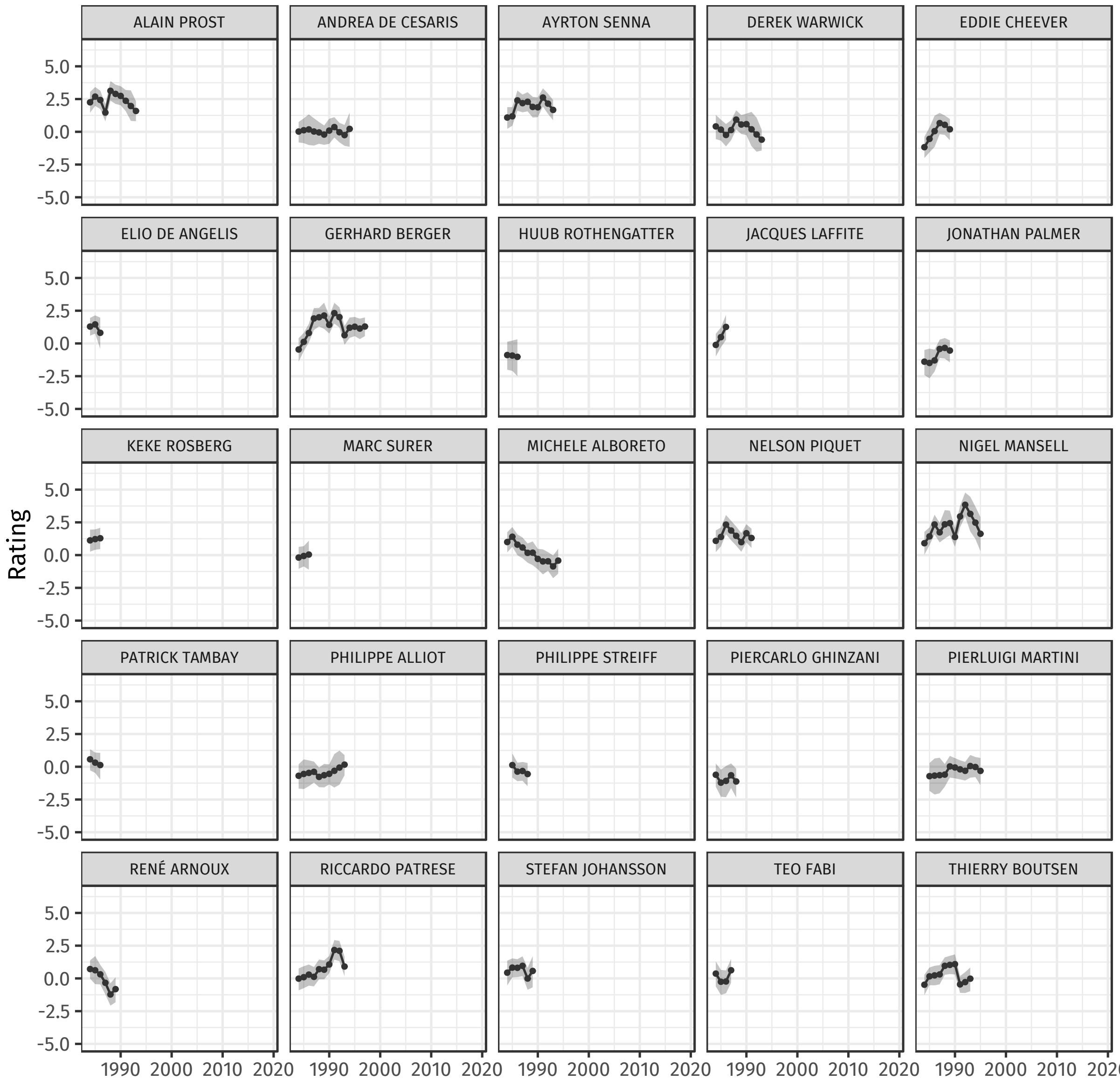


Figure A.3: Estimated Dynamic Ratings. Only showing drivers that have more than three years of career. Solid lines are posterior medians and shaded areas are 95% credible intervals based on the posterior quantile. The rating is relative to Timo Glock's first year (2004) which is fixed to zero for identification. Missing years are interpolated via predicted values.

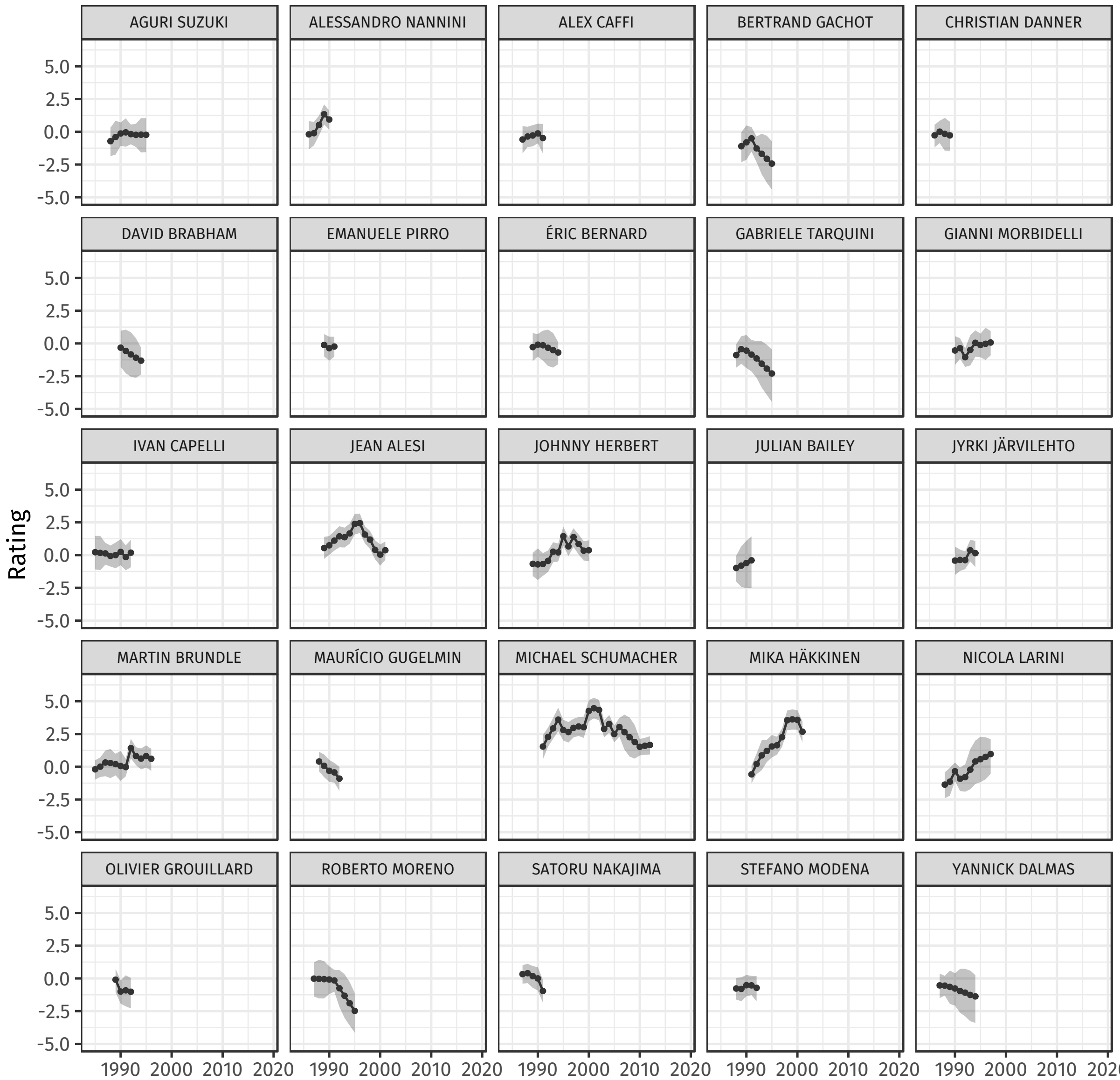


Figure A.4: Estimated Dynamic Ratings (Cont'). Only showing drivers that have more than three years of career. Solid lines are posterior medians and shaded areas are 95% credible intervals based on the posterior quantile. The rating is relative to Timo Glock's first year (2004) which is fixed to zero for identification. Missing years are interpolated via predicted values.

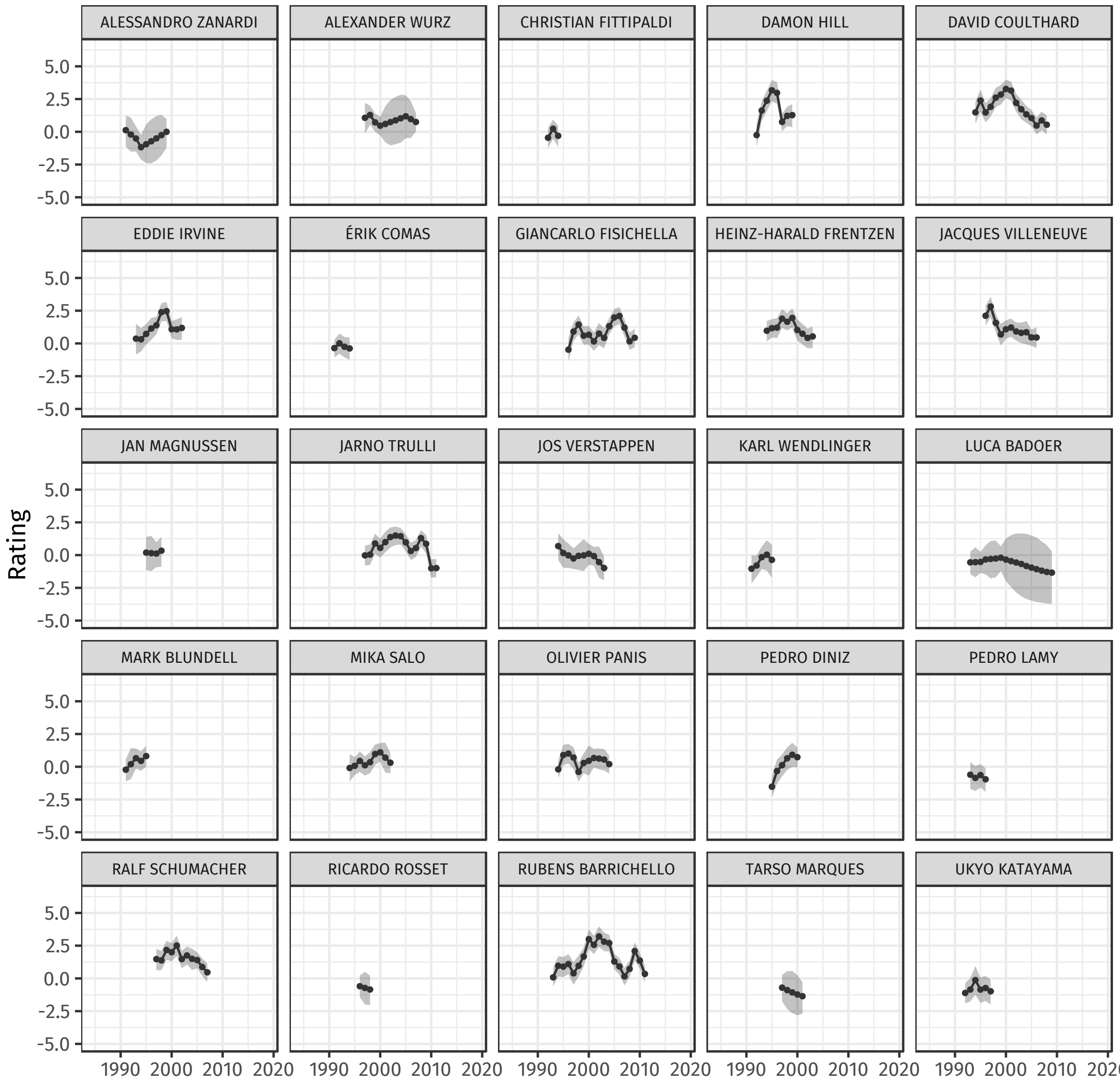


Figure A.5: Estimated Dynamic Ratings (Cont'). Only showing drivers that have more than three years of career. Solid lines are posterior medians and shaded areas are 95% credible intervals based on the posterior quantile. The rating is relative to Timo Glock's first year (2004) which is fixed to zero for identification. Missing years are interpolated via predicted values.

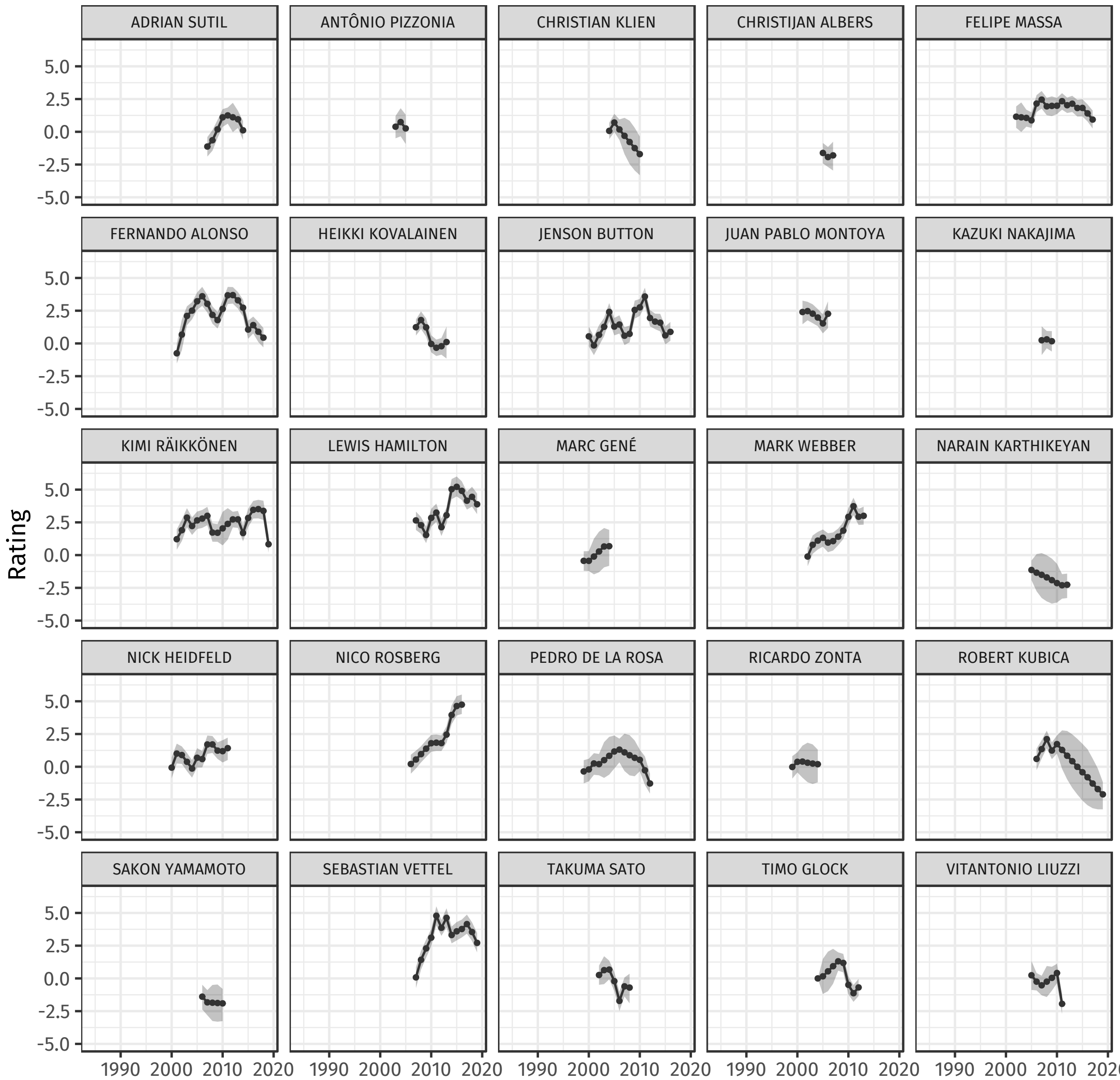


Figure A.6: Estimated Dynamic Ratings (Cont'). Only showing drivers that have more than three years of career. Solid lines are posterior medians and shaded areas are 95% credible intervals based on the posterior quantile. The rating is relative to Timo Glock's first year (2004) which is fixed to zero for identification. Missing years are interpolated via predicted values.

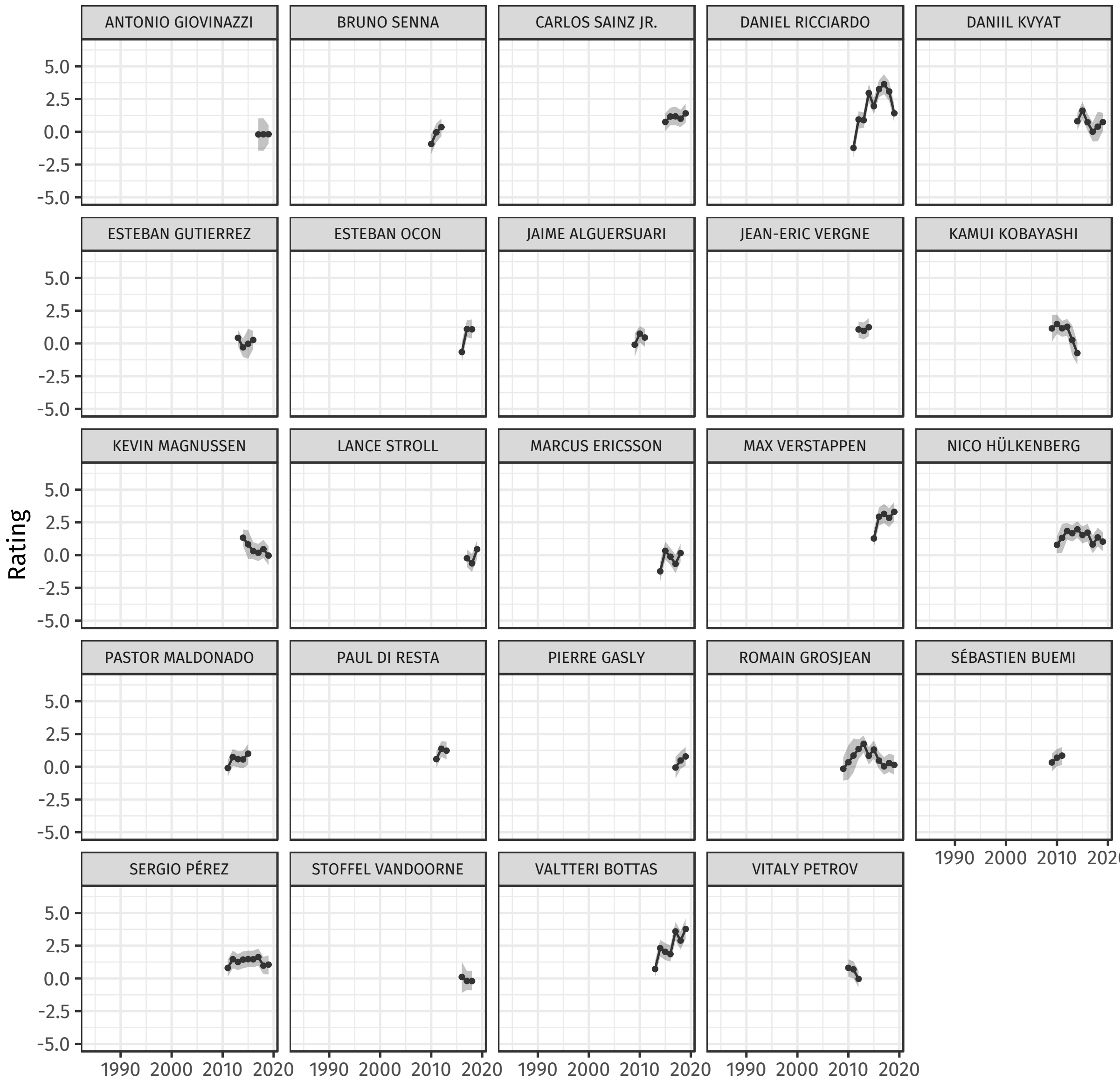


Figure A.7: Estimated Dynamic Ratings (Cont'). Only showing drivers that have more than three years of career. Solid lines are posterior medians and shaded areas are 95% credible intervals based on the posterior quantile. The rating is relative to Timo Glock's first year (2004) which is fixed to zero for identification. Missing years are interpolated via predicted values.

B Additional Results: Team's Rating

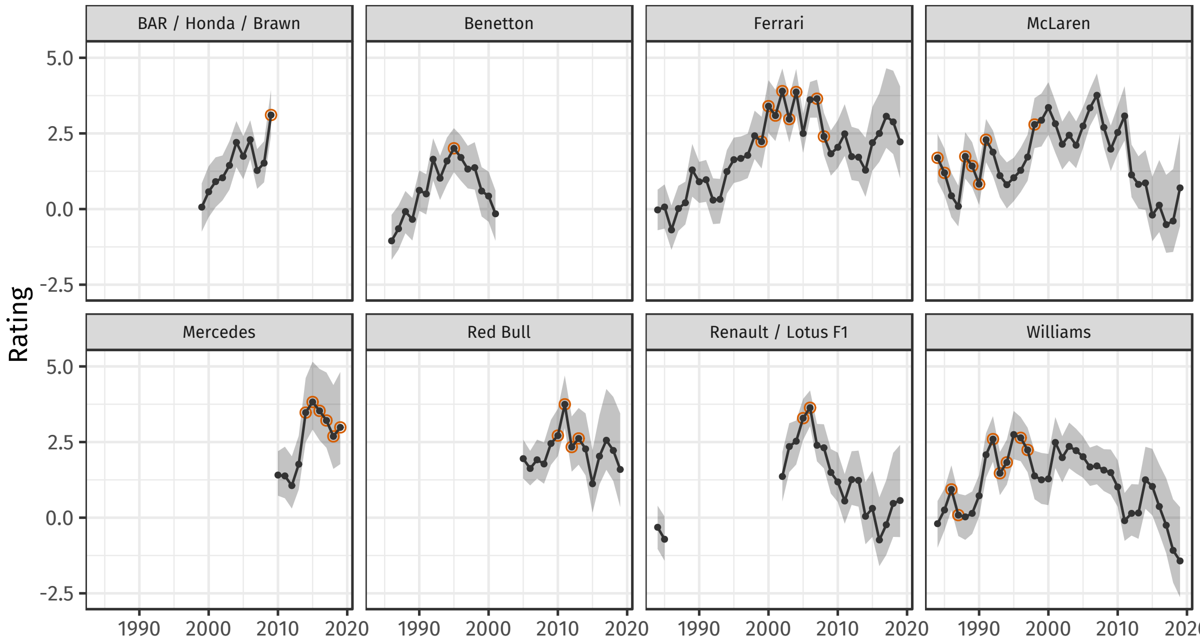


Figure B.1: Dynamic Rating Estimates for World Champion Constructors. Solid orange lines represent the year of the championship. Estimates are based on the entire teams in the data (1984–2019). The rating is relative to the first year of Super Aguri (2006), which is fixed to 0. Solid lines in black show the posterior medians and gray areas show the quantile-based 95% credible intervals.

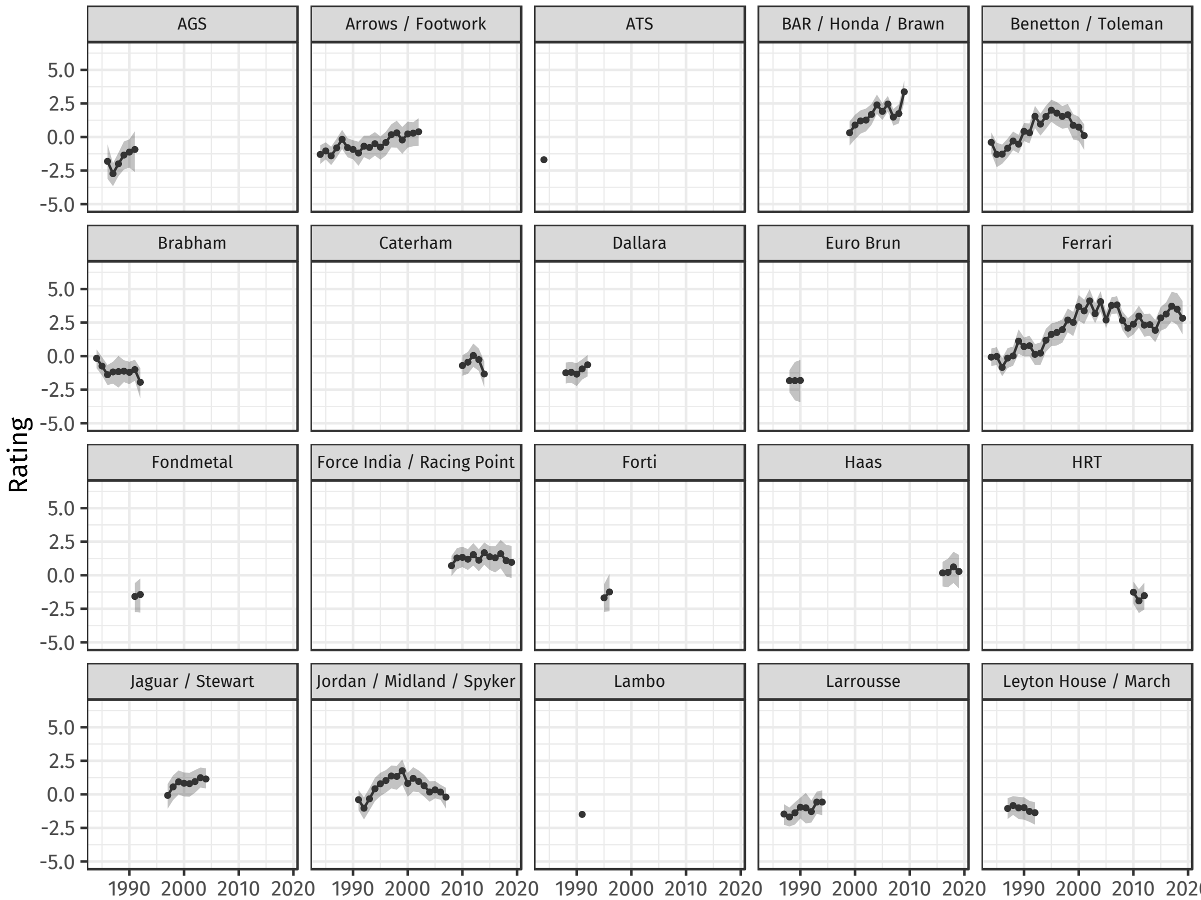


Figure B.2: Dynamic Rating Estimates for Formula One Teams. Presented are teams with more than two years in the racing. Estimates are based on the entire teams in the data (1984–2019). The rating is relative to the first year of Super Aguri (2006), which is fixed to 0. Solid lines show the posterior medians and gray areas show the quantile-based 95% credible intervals.

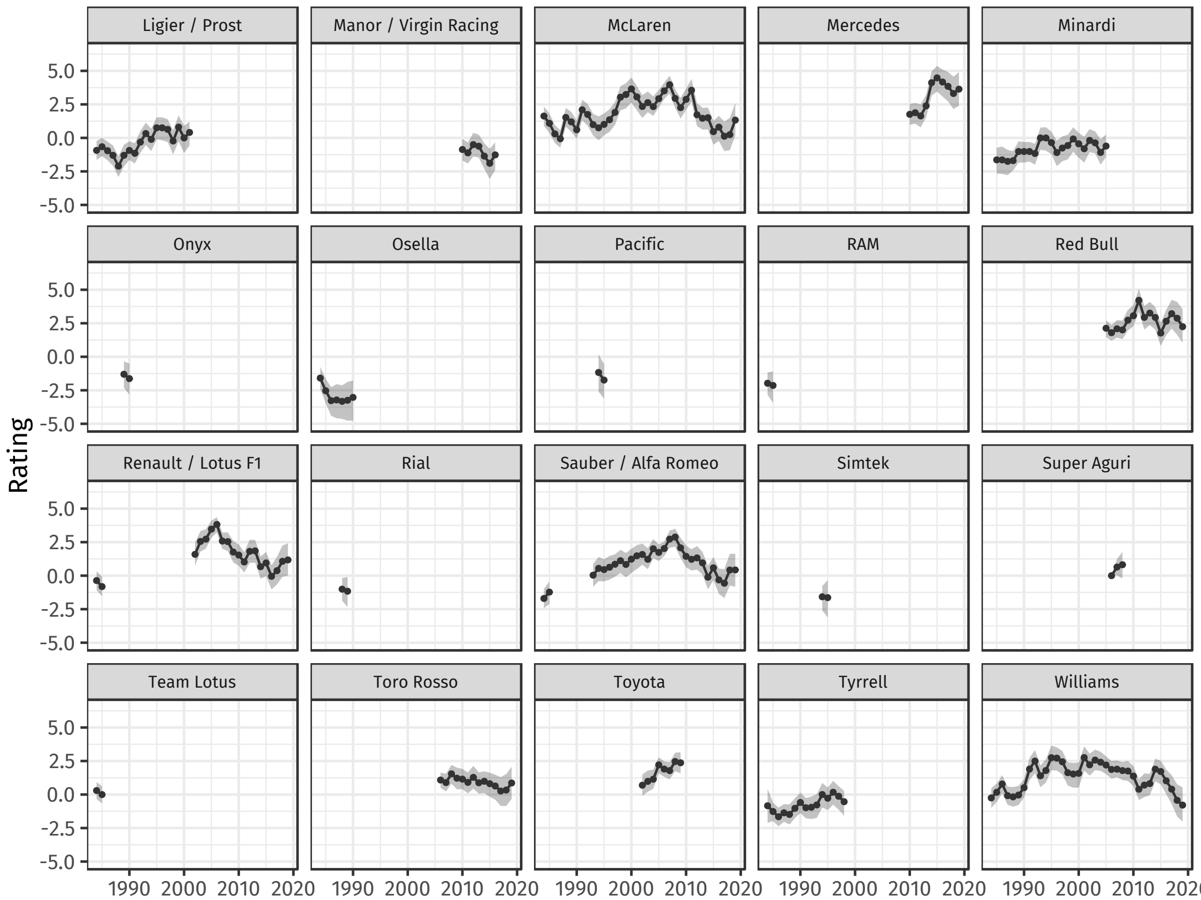


Figure B.3: Dynamic Rating Estimates for Formula One Teams (Cont'). Presented are teams with more than two years in the racing. Estimates are based on the entire teams in the data (1984–2019). The rating is relative to the first year of Super Aguri (2006), which is fixed to 0. Solid lines show the posterior medians and gray areas show the quantile-based 95% credible intervals.