Modelling momentum in football

Marius Ötting[1], Roland Langrock[1], Antonello Maruotti[2]

[1]Bielefeld University, Germany

[2]Libera Universita Maria Ss. Assunta, Italy

July 1, 2019

MathSport International, Athens

@marius_oetting

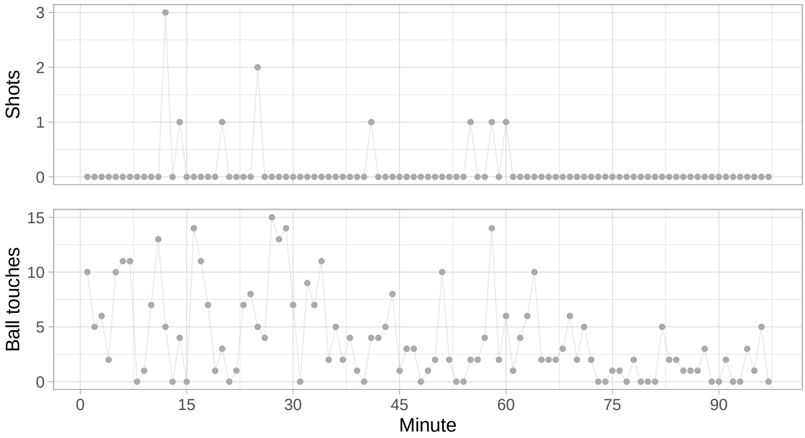
But Bernardo Silva's deflected effort and Sterling's second saw the momentum switch back to the Etihad side who then only needed one more.

"Soccer is a game of momentum, and you definitely felt the momentum switch there," Stone said of his second-half header. "But it just didn't turn out the way we wanted."

"I think getting the two goals very quickly at the end of the half was the key, the momentum switch."

Data

vs. (November 25, 2017)

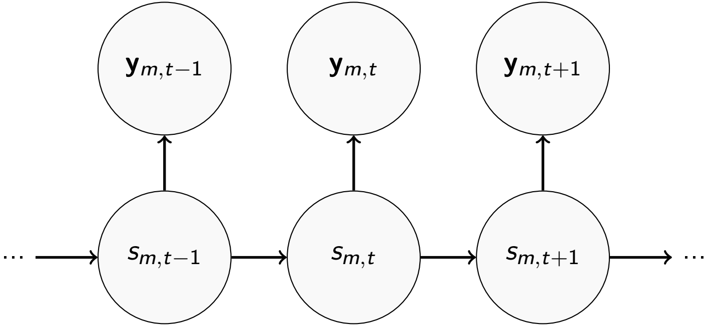


Data

- minute-by-minute data taken from `www.whoscored.com`
- multivariate time series $\{\mathbf{y}_{mt}\}_{t=1,2,\ldots,T_m}$
- $\mathbf{y}_{mt} = (y_{mt1}, \ldots, y_{mtK})$: obs. variables in match m at time t
- here, $K = 2$ variables: **shots on goal** and **ball touches**

Modelling momentum

- a match progresses through different phases
  $\rightsquigarrow$ hidden Markov models (HMMs) are considered



- observations $\mathbf{y}_{mt}$ are driven by a state process s_{mt}
- transition probability matrix (t.p.m.) $\mathbf{\Gamma} = (\gamma_{ij})$ for s_{mt}

State-dependent distributions

- shots on goal and ball touches are count variables
  (hence Poisson distribtion would be a standard choice)

- here, we account for possible over- and underdispersion
  $\rightsquigarrow$ Conway-Maxwell-Poisson distribution is considered

$$\Pr(X = x) = \frac{1}{Z(\lambda, \nu)} \frac{\lambda^x}{(x!)^\nu}$$

  with $Z(\lambda, \nu) = \sum_{k=0}^{\infty} \lambda^k / (k!)^\nu$, $\lambda > 0$ and $\nu \geq 0$.

State-dependent distributions

- within-state correlation in $\mathbf{y}_{mt}$ is allowed by using a copula C
- bivariate distribution as state-dependent distribution:

$$F(\mathbf{y}_{mt} \mid s_{mt}) = C\big(F_1(y_{mt1} \mid s_{mt}), F_2(y_{mt2} \mid s_{mt})\big)$$

  - F_1, F_2: c.d.f. of the two marginals
  - C: copula

State-dependent distributions

- differences have to be taken for the joint p.m.f. (discrete marginals!) (see, e.g., Nikoloulopoulos 2013):

$$\begin{aligned} f(\mathbf{y}_{mt} \,|\, s_{mt}) = {} & C\big(F_1(y_{mt1} \,|\, s_{mt}), F_2(y_{mt2} \,|\, s_{mt})\big) \\ & - C\big(F_1(y_{mt1} - 1 \,|\, s_{mt}), F_2(y_{mt2} \,|\, s_{mt})\big) \\ & - C\big(F_1(y_{mt1} \,|\, s_{mt}), F_2(y_{mt2} - 1 \,|\, s_{mt})\big) \\ & + C\big(F_1(y_{mt1} - 1 \,|\, s_{mt}), F_2(y_{mt2} - 1 \,|\, s_{mt})\big) \end{aligned}$$

- copulas allowing for positive and negative dep. are considered (Frank, Ali-Mikhail-Haq (AMH) and Clayton)

Transistion probabilities

- modeling state-switching by covariates
  $\rightsquigarrow$ entries $\gamma_{ij}^{(t)}$ in the t.p.m. are functions of covariates

- covariates are contained in linear predictor $\eta_{ij}^{(t)}$:

$$\gamma_{ij}^{(t)} = \frac{\exp(\eta_{ij}^{(t)})}{1 + \sum_{l \neq i} \exp(\eta_{il}^{(t)})}$$

- t.p.m. $\mathbf{\Gamma}_t$ is not constant anymore

HMM likelihood

With the t.p.m. $\boldsymbol{\Gamma}_t$ as above and..

- ... $N \times N$ diagonal matrix $\mathbf{P}(\mathbf{y}_{mt})$ with i-th diagonal element given by $f(\mathbf{y}_{mt} | s_{mt} = i)$
- ... $\boldsymbol{\delta} = \big(\Pr(s_{m1} = 1), \dots, \Pr(s_{m1} = N)\big)$

likelihood for a single match m given as: (see Zucchini et al. 2016)

$$L = \boldsymbol{\delta} \mathbf{P}(\mathbf{y}_{m1}) \boldsymbol{\Gamma}_t \mathbf{P}(\mathbf{y}_{m2}) \dots \boldsymbol{\Gamma}_t \mathbf{P}(\mathbf{y}_{mT_m}) \mathbf{1}$$

(with $\mathbf{1} = (1, \dots, 1)' \in \mathbb{R}^N$)

Results

Results

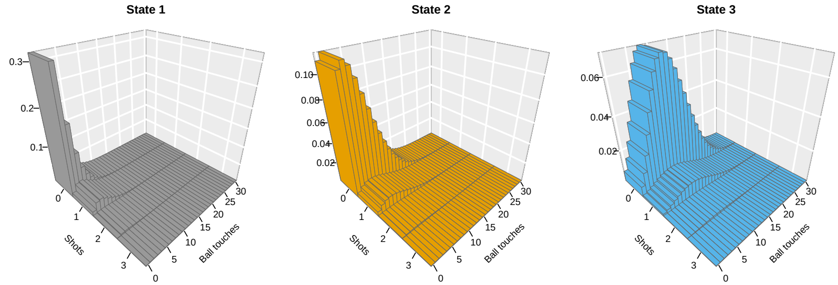
- data of Borussia Dortmund (season 2017/18) ($\rightsquigarrow$ $m = 34$ matches)
- covariates:
  - difference in the current score
  - market value of the opponent (in Mio. Euro)
  - minute of the match

Results – model selection

model selection by AIC and BIC:

	Frank		Clayton		AMH	
	AIC	BIC	AIC	BIC	AIC	BIC
2 states	20,954	21,033	20,941	21,020	20,943	21,022
3 states	20,865	**21,005**	20,839	**20,979**	20,861	**21,001**
4 states	20,836	21,049	20,817	21,030	**20,831**	21,043
5 states	**20,814**	21,112	**20,801**	21,098	20,834	21,132

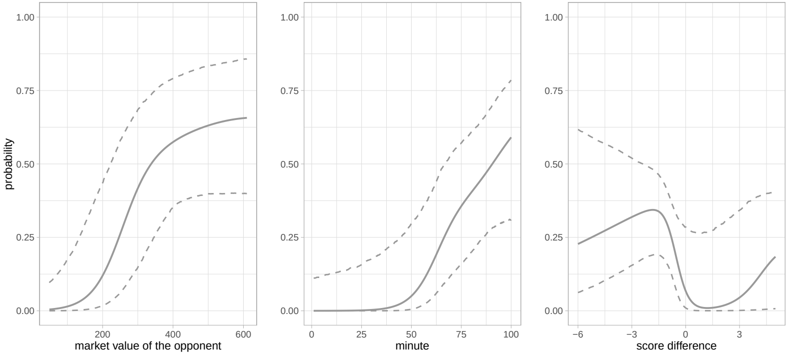
Results – state-dependent distributions



Shots on goal	$\hat{\lambda} = 0.0001, \hat{\nu} = 1.507$ mean: 0.148	$\hat{\lambda} = 0.132, \nu = 0.256$ mean: 0.133	$\hat{\lambda} = 0.150, \hat{\nu} = 0.071$ mean: 0.173
Ball touches	$\hat{\lambda} = 0.131, \hat{\nu} = 0.003$ mean: 2.368	$\hat{\lambda} = 1.077, \hat{\nu} = 0.164$ mean: 4.772	$\hat{\lambda} = 1.638, \hat{\nu} = 0.256$ mean: 9.947
Dependence	$\hat{\theta} = 2.241$	$\hat{\theta} = 0.225$	$\hat{\theta} = -0.226$

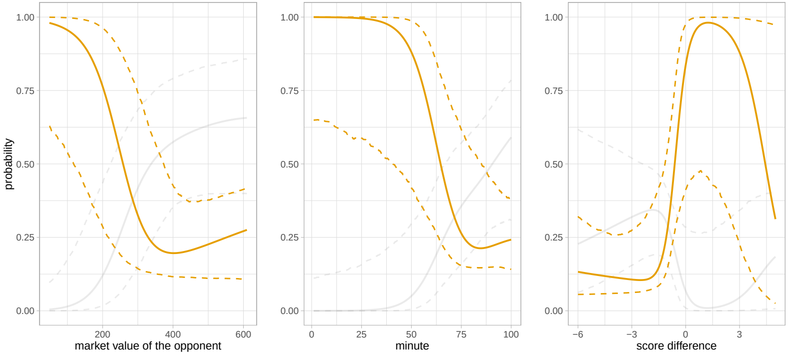
Results – stationary distributions

State 1: defense and counter attacks; State 2: balanced; State 3: dominance



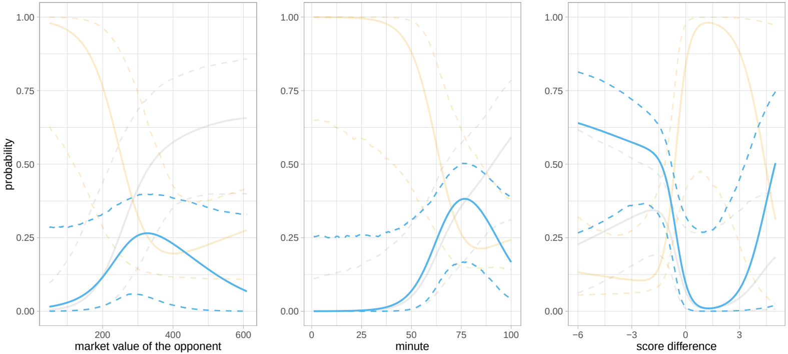
Results – stationary distributions

State 1: defense and counter attacks; State 2: balanced; State 3: dominance



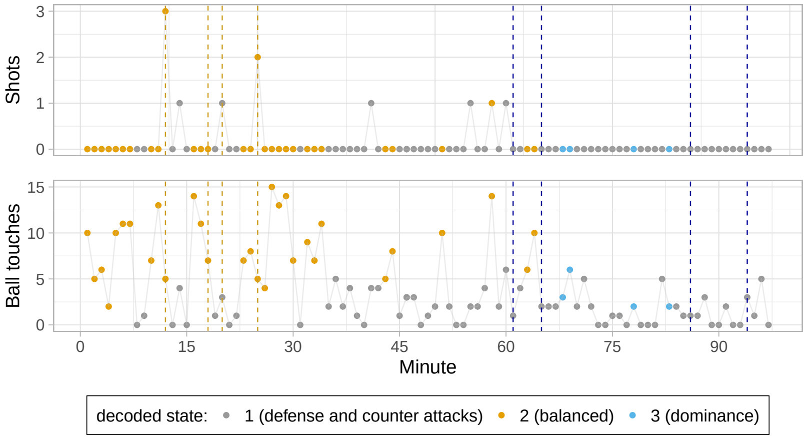
Results – stationary distributions

State 1: defense and counter attacks; State 2: balanced; State 3: dominance



Results: decoded state sequence

vs. (November 25, 2017)



Conclusions & current research

- model potentially useful for managers but also for bookmakers
  - analysis of opponents
  - predicting goals for in-game betting
- dealing with short interruptions

References

Nikoloulopoulos, A. K. (2013). Copula-based models for multivariate discrete response data. In Jaworski, P., Durante, F., and Härdle, W. K., editors, *Copulae in Mathematical and Quantitative Finance*, pages 231–249. Heidelberg: Springer.

Zucchini, W., MacDonald, I. L., and Langrock, R. (2016). *Hidden Markov Models for Time Series: An Introduction Using R*. Boca Raton: Chapman & Hall/CRC.