**The Role of Marketer Generated Content in Customer Engagement Marketing**

**(JM.18.0306.R4)**

**WEB APPENDIX**

**Appendix W1: Event Outcomes and Related Concepts in the Literature**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Construct** | **Citation** | **Conceptualization** | **Objective** | **Single “event”** | **Operationalization** |
| ***Event outcomes*** | ***This study*** | Observed level of firm performance during customer brand- or firm-related interaction | Y | Y | Win/draw/lose outcome *for a single match*, with model also accounting for odds published prior to match, stakes involved, and opponent quality |
| ***Service quality*** | Nam, Manchanda, and Chintagunta (2010) | Objective service performance | Y | N | Video-on-Demand signal quality aggregated to the weekly level |
| Gijsenberg, van Heerde and Verhoef (2015) | Objective service performance | Y | N | Percentage of successful railway connections (train departure is more than the minimum required to make a connection from a prior ride), aggregated to the monthly level |
| ***Satisfaction*** | Hart, Kerrigan, vom Lehn (2016) | Consumption-related fulfillment | N | Y | Coding of introspective diary reports on experiences watching a film |
| Farhadloo, Patterson, and Rolland (2016) | Customer satisfaction with a product, service, or experience | N | N | Probabilistic model enabling the discovery of the relative importance of different features for unique products, services, or experiences |
| ***Brand evaluations*** | Beukeboom, Kerkhof, and de Vries (2015) | Brand evaluations based on interactions | N | N | Attitudinal survey measures of willingness to recommend |

**Appendix W2. Details on Unexpected Outcomes**

We follow an approach similar to previous research using odds to account for customer expectations (Bartling, Brandes, and Schunk 2015; Card and Dahl 2011). More specifically, we gather, for each of the matches, the pre-play betting odds for each possible outcome (win, loss, or draw, seen from the point of view of the focal team)[[1]](#footnote-2). We compare the actual match result with expectations (odds) and identify unexpected outcomes. Next, we convert the betting odds into probabilities for each outcome by taking the inverse of these odds (the odds are displayed as decimals). Since the odds include a bookmaker’s margin, the probabilities sum up to more than one, so we rescale the probabilities to one (Bartling, Brandes, and Schunk 2015). As an example, the win, draw, and losing odds for the focal team during one match were 1.40, 4.78, and 7.66 respectively. Thus, a one euro bet on a focal team win would result in receiving 1.40 euros in the case the focal team wins (a gain of .40 euros), and otherwise losing one euro. The probability of winning for the home team in this case is 67.77%[[2]](#footnote-3). Next, we observe that a draw is never the expected game outcome, i.e., the one with the highest probability[[3]](#footnote-4). We therefore rescale the variable to represent the probability of the focal team winning versus the probability of the focal team losing (Dolton and Mackerron 2018). If the probability is higher (lower) than .50, the expectation is that the game will be won (lost). Finally, in line with research including sports bets (Bartling, Brandes, and Schunk 2015; Card and Dahl 2011; Dolton and Mackerron 2018), we compare the actual match result with expectations (odds) and identify unexpected results. In case the actual result was a win and a loss was expected (probability of a win < 50%), we call this an unexpected or upset win. Note that in the case of a draw, we consider this to be expected if the probability of a win was between 40% and 60%; that is, the probability of winning and losing are relatively close[[4]](#footnote-5) (Card and Dahl 2011). We include the binary variable Unexpected Result in our analyses.

**Appendix W3. Descriptives for the Selection Equation and Endogeneity Control Functions**

**Appendix W3.1: Selection Equation**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***Customer comment Selection Equation*** | |  | |  |  |  |
| *Dependent variable* |  | |  | M | SD | Range |
| *Commenting* | Dummy variable indicating whether a customer *u* has commented over the timeframe of the study (2011-2014) | |  | .18 | .38 | [0,1] |
| *Explanatory variables* |  | |  |  |  |  |
| *Age* | Age of the customer | |  | 41.74 | 15.51 | [11.65,92.51] |
| *Gender* | Gender of the customer (1 = male; 0 = female) | |  | .92 | .27 | [0,1] |
| *Language* | Dummy variable indicating the language of the customer (1= Dutch; 0 = other) | |  | .97 | .18 | (0,1] |
| *OnlinePurchase* | Number of tickets purchased online in the past | |  | .65 | 1.65 | [0,36] |
| *Recency t-1* | Recency (in years) of the last purchase of a season ticket amount (at time *t-1*) | |  | 1.78 | 1.82 | [0,10] |
| *Tenure t-1* | Length of relationship (in years) of the focal customer with the company at time *t*-1 | |  | 6.20 | 3.43 | [1,11] |

**Appendix W3.2: Endogeneity Control Functions**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Endogeneity correction regressions*** | |  | | |  | |  | |  | |
| *Dependent variables* |  | |  | M | | SD | | Range | |
| *InformationalMGCp,m* | Number of informational posts on Facebook by the focal team between the end of match *m* and the time of publishing post *p* | |  | 3.83 | | 6.13 | | [0,49] | |
| *EmotionalMGCp,m* | Number of emotional posts on Facebook by the focal team between the end of match *m* and the time of publishing post *p* | |  | 3.29 | | 6.56 | | [0,47] | |
| *Explanatory variables* |  | |  |  | |  | |  | |
| *Unexpected Resultm* | Dummy variable indicating whether the result of match *m* is in line with the expectations based on the odds | |  | .33 | | .47 | | [0,1] | |
| *Result m \* Unexpected Result m* | Interaction effect between actual Result of match *m* and the dummy variable *Unexpected Result* | |  |  | |  | |  | |
| *Event Attendancem* | The number of spectators for game *m*, which is a proxy for the importance of the game and quality of the opponent | |  | 18,982 | | 8,941 | | [1,819;65,110] | |
| *Result m \* Event Attendance m* | Interaction effect between actual Result of match *m* and the dummy variable *Spectators* | |  |  | |  | |  | |
| *RedCardsm* | The number of red cards for the focal team in match *m* | |  | .14 | | .36 | | [0,1] | |
| *YellowCardsm* | The number of yelllow cards for the focal team in match *m* | |  | 1.88 | | 1.24 | | [0,6] | |
| *Home Matchm* | Dummy indicating whether the match *m* is a home match | |  | .55 | | .5 | | [0,1] | |
| *Post Timep,m* | Time (in hours) that has passed by (at the moment of publishing a post *p* by the team) since the end of the match *m* | |  | 18.26 | | 16.22 | | [.04,46.16] | |
| *Δ Informational postsm* | The difference in volume of informational posts by the team in the two months before the match *m* (volume in month-1 before match – volume in month-2 before match) | |  | -.79 | | 29.70 | | [-81,111] | |
| *Δ Emotional postsm* | The difference in volume of emotional posts by the team in the two months before the match *m* (volume in month-1 before match – volume in month-2 before match) | |  | .10 | | 27.16 | | [-118,131] | |

**Appendix W4. Results of Multinomial Regression Model (Negative vs Neutral)**

**Appendix W4.1: Results**

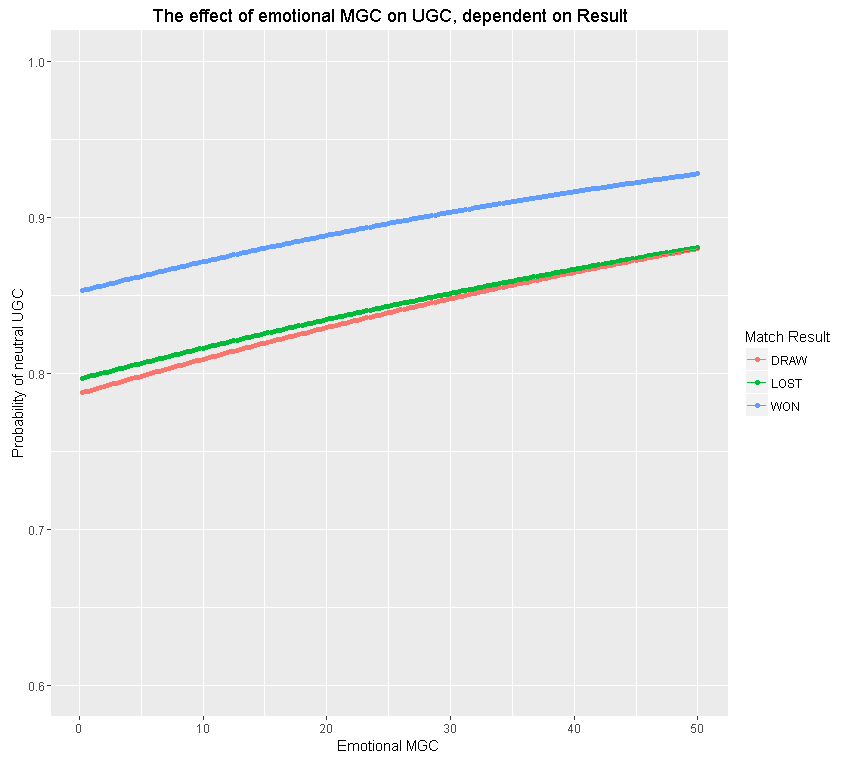
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variables | Estimate | | z-score |  |
| *Intercept* | 1.355 | \*\*\* | 11.004 |  |
| *Result (Lost)m* | .057 |  | .349 |  |
| *Result (Won)m* | .447 | \*\*\* | 3.910 |  |
| *Informational MGCu,c,m* | .061 |  | .991 |  |
| *ResultLostm \* Informational MGCu,c,m* | .623 | \*\*\* | 3.022 |  |
| *ResultWonm \* Informational MGCu,c,m* | -.035 |  | -.495 |  |
| *Emotional MGCu,c,m* | .171 |  | 1.532 |  |
| *ResultLostm \* Emotional MGCu,c,m* | -.012 |  | -.091 |  |
| *ResultWonm \* Emotional MGCu,c,m* | .029 |  | .254 |  |
| *Unexpected Resultm* | .033 |  | .251 |  |
| *ResultLostm \* Unexpected m* | .108 |  | .576 |  |
| *ResultWonm \* Unexpected m* | .001 |  | .005 |  |
| *TotalEventAttendancem* | -.027 |  | -.350 |  |
| *ResultLostm \* TotalEventAttendance m* | .072 |  | .738 |  |
| *ResultWonm \* TotalEventAttendance m* | .067 |  | .728 |  |
| *RedCardsm* | -.034 |  | -.987 |  |
| *YellowCardsm* | -.015 |  | -.434 |  |
| *Home Game m* | -.123 | \* | -1.704 |  |
| *EventFacebooku,m* | .174 |  | 1.512 |  |
| *Customer Sentiment u,c,m-1* | .041 |  | .756 |  |
| *Other Sentiment Valence u,c,m* | .020 |  | .864 |  |
| *Other Sentiment Volume u,c,m* | .061 | \*\* | 2.168 |  |
| *Comment length u,c,m* | -.967 | \*\*\* | -35.649 |  |
| *Comment time u,c,m* | .189 | \*\*\* | 3.379 |  |
| *ResultLost m \* Comment time u,c,m* | .080 |  | 1.068 |  |
| *ResultWon m \* Comment time u,c,m* | -.151 | \*\* | -2.067 |  |
| *IMRu* | .004 |  | .174 |  |
| *Endogeneity Correction Informational MGC m* | -.070 |  | -1.335 |  |
| *Endogeneity Correction Emotional MGC m* | -.001 |  | -.015 |  |
| *N (observations)\** | 21,604 | | |  |

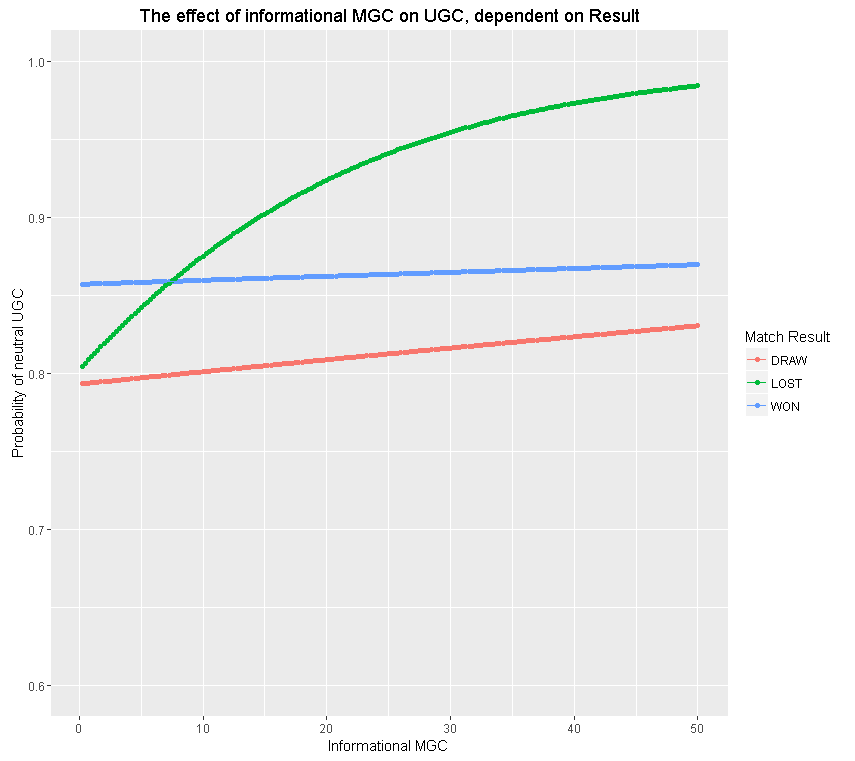
Note: \* *p*<.1, \*\* *p*<.05, \*\*\* *p*<.01; coefficients are standardized.

\* Note that the approach followed to allows to include all posts, not only the posts categorized as positive or negative by the dictionary-based approach used in the main analysis.

We do not repeat the results for the negative vs positive model, since this is essentially the same as our main sentiment equation. There are two differences: first, in a multinomial model we estimate it jointly with the negative vs neutral model. Second, the standardization is done on the entire set of positive, negative and neutral comments, making it impossible to directly compare these coefficients with the ones of the main model (which are standardized on only positive and negative comments).

**Appendix W4.2: Interaction Plots with MGC**





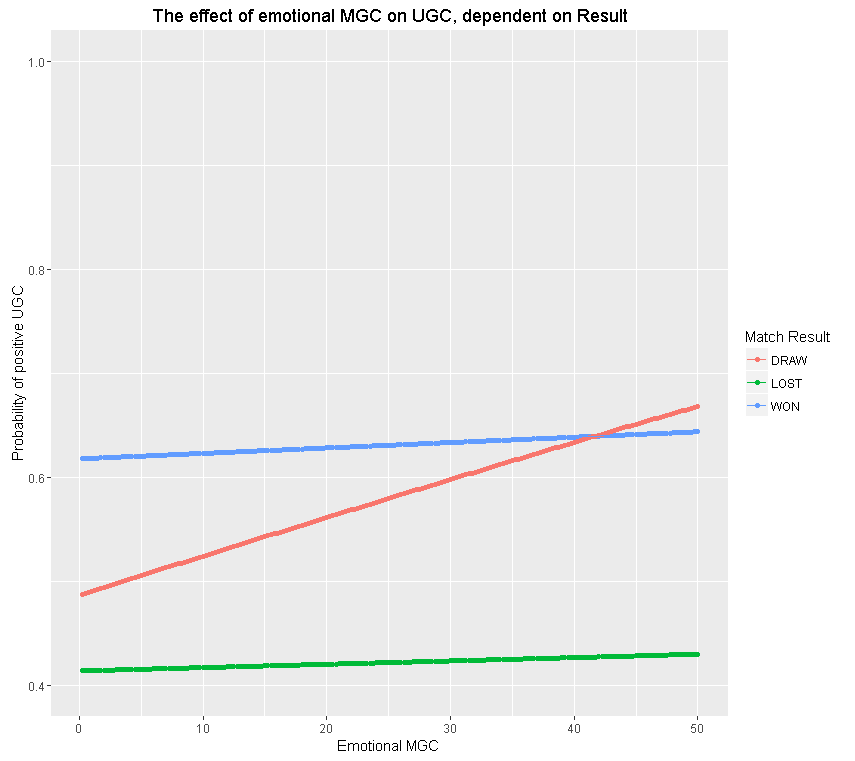
**Appendix W5. Results of Fractional Regression Model**

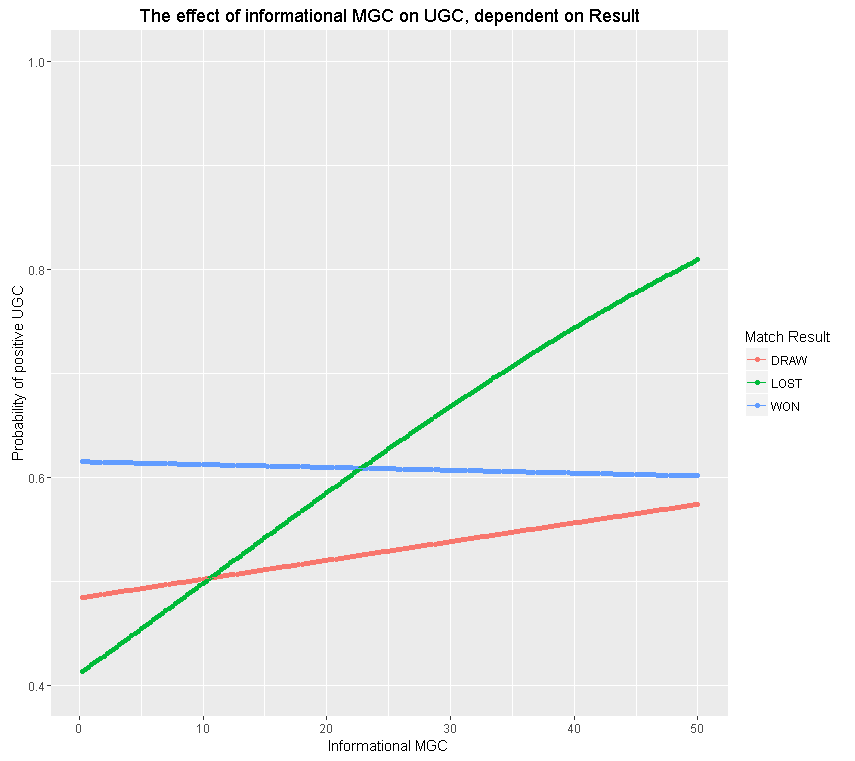
**Appendix W5.1: Results**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variables | Estimate | | z-score |  |
| *Intercept* | .103 | \* | 1.704 |  |
| *Result (Lost)m* | -.008 |  | -.103 |  |
| *Result (Won)m* | .193 | \*\*\* | 3.792 |  |
| *Informational MGCu,c,m* | .019 |  | 1.064 |  |
| *ResultLostm \* Informational MGCu,c,m* | .072 | \*\* | 2.092 |  |
| *ResultWonm \* Informational MGCu,c,m* | -.022 |  | -.981 |  |
| *Emotional MGCu,c,m* | .065 | \*\*\* | 2.647 |  |
| *ResultLostm \* Emotional MGCu,c,m* | -.060 | \* | -1.85 |  |
| *ResultWonm \* Emotional MGCu,c,m* | -.056 | \*\* | -2.377 |  |
| *Unexpected Resultm* | -.025 |  | -.411 |  |
| *ResultLostm \* Unexpected m* | -.068 |  | -.770 |  |
| *ResultWonm \* Unexpected m* | .201 | \*\*\* | 2.621 |  |
| *TotalEventAttendancem* | -.047 |  | -1.623 |  |
| *ResultLostm \* TotalEventAttendance m* | .050 |  | 1.287 |  |
| *ResultWonm \* TotalEventAttendance m* | .066 | \*\* | 1.986 |  |
| *RedCardsm* | -.016 |  | -.968 |  |
| *YellowCardsm* | -.019 |  | -1.306 |  |
| *Home Game m* | -.008 |  | -.264 |  |
| *EventFacebooku,m* | -.004 |  | -.152 |  |
| *Customer Sentiment u,c,m-1* | .045 | \*\*\* | 4.078 |  |
| *Other Sentiment Valence u,c,m* | .030 | \*\*\* | 5.801 |  |
| *Other Sentiment Volume u,c,m* | .028 |  | 1.463 |  |
| *Comment length u,c,m* | -.192 | \*\*\* | -19.632 |  |
| *Comment time u,c,m* | .107 | \*\*\* | 3.431 |  |
| *ResultLost m \* Comment time u,c,m* | .052 |  | 1.357 |  |
| *ResultWon m \* Comment time u,c,m* | -.144 | \*\*\* | -3.976 |  |
| *IMRu* | .014 | \*\* | 2.385 |  |
| *Endogeneity Correction Informational MGC m* | -.048 | \*\* | -2.199 |  |
| *Endogeneity Correction Emotional MGC m* | -.021 |  | -1.596 |  |
|  |  | | |  |
| *N (observations)\** | 21,604 | | |  |

\* Note that the approach followed to allows to include all posts, not only the posts categorized as positive or negative by the dictionary-based approach used in the main analysis.

**Appendix W5.2: Interaction Plots**





**Appendix W6. Results of the Customer Sentiment Model with Squared Volume of Informational and Emotional MGC**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variables | Estimate | | z-score |  |
| *Intercept* | .717 | \*\*\* | 9.221 |  |
| *Result (Lost)m* | -.101 |  | -.975 |  |
| *Result (Won)m* | .252 | \*\*\* | 3.483 |  |
| *Informational MGCu,c,m* | .044 |  | 1.044 |  |
| *ResultLostm \* Informational MGCu,c,m* | .344 | \*\*\* | 2.626 |  |
| *ResultWonm \* Informational MGC u,c,m* | -.021 |  | -.417 |  |
| *Squared Informational MGCu,c,m* | .026 |  | .786 |  |
| *ResultLostm \* Squared Informational MGCu,c,m* | -.028 |  | -.470 |  |
| *ResultWonm \* Squared Informational MGC u,c,m* | -.023 |  | .035 |  |
| *Emotional MGC u,c,m* | .179 | \*\*\* | 2.776 |  |
| *ResultLostm \* Emotional MGCu,c,m* | -.033 |  | -.425 |  |
| *ResultWonm \* Emotional MGC u,c,m* | -.105 |  | -1.635 |  |
| *Squared Emotional MGC u,c,m* | -.147 | \*\* | -2.379 |  |
| *ResultLostm \* Squared Emotional MGCu,c,m* | .043 |  | .526 |  |
| *ResultWonm \* Squared Emotional MGC u,c,m* | .157 | \*\* | 2.429 |  |
| *Unexpected Resultm* | -.095 |  | -1.098 |  |
| *ResultLostm \* Unexpected m* | .057 |  | .453 |  |
| *ResultWonm \* Unexpected m* | .265 | \*\* | .035 |  |
| *TotalEventAttendancem* | -.115 | \*\* | -2.349 |  |
| *ResultLostm \* TotalEventAttendance m* | .091 |  | 1.411 |  |
| *ResultWonm \* TotalEventAttendance m* | .145 | \*\* | .035 |  |
| *RedCardsm* | -.044 | \* | -1.880 |  |
| *YellowCardsm* | -.021 |  | -.959 |  |
| *Home Game m* | -.050 |  | -1.074 |  |
| *EventFacebooku,m* | .042 |  | .626 |  |
| *Customer Sentiment u,c,m-1* | .086 | \*\*\* | 2.759 |  |
| *Other Sentiment Valence u,c,m* | .046 | \*\*\* | 3.159 |  |
| *Other Sentiment Volume u,c,m* | .042 | \*\* | 2.354 |  |
| *Comment length u,c,m* | -.073 | \*\*\* | -4.742 |  |
| *Comment time u,c,m* | .154 | \*\*\* | 4.523 |  |
| *ResultLost m \* Comment time u,c,m* | .091 | \*\* | 2.011 |  |
| *ResultWon m \* Comment time u,c,m* | -.152 | \*\*\* | -3.501 |  |
| *IMRu* | .028 | \* | 1.842 |  |
| *Endogeneity Correction Informational MGC m* | -.070 | \*\* | -2.072 |  |
| *Endogeneity Correction Emotional MGC m* | -.041 | \* | -1.816 |  |
|  |  | | |  |
| *Log-Likelihood* | -5,580.6 | | |  |
| *AIC* | 11,243.25 | | |  |

Note: \* *p*<.1, \*\* *p*<.05, \*\*\* *p*<.01; coefficients are standardized. The standard errors are bootstrapped.

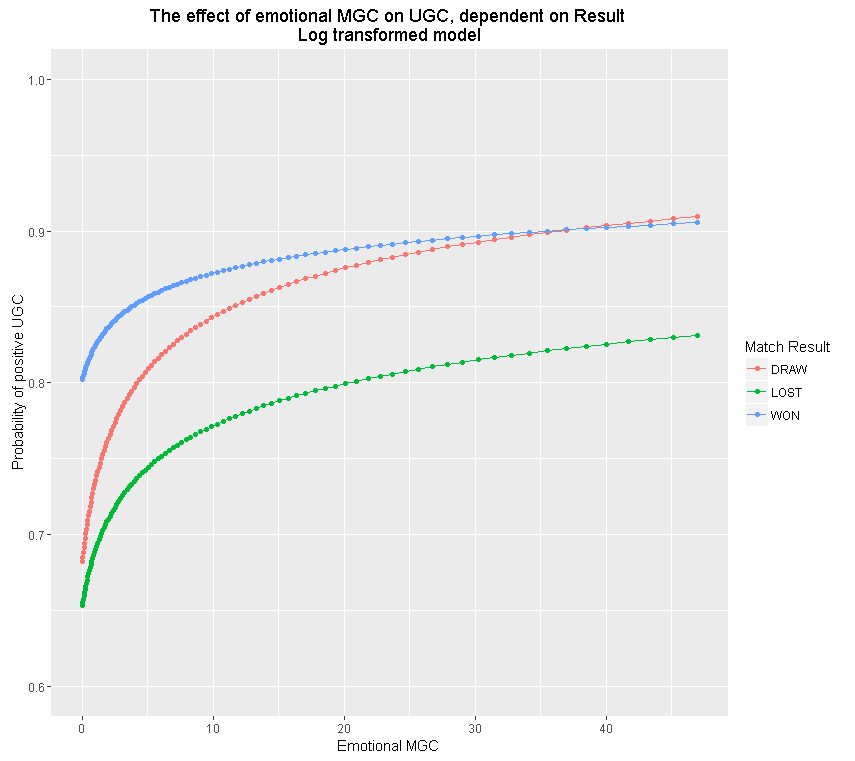
**Appendix W7: Results of the Customer Sentiment Model with Log-transformed Volume of Informational and Emotional MGC**

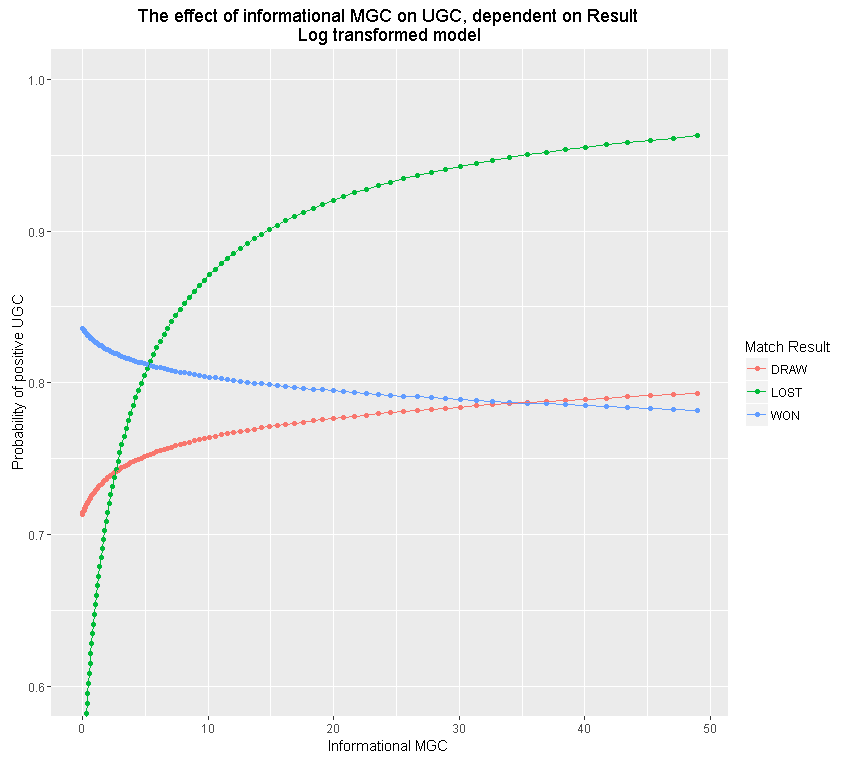
**Appendix W7.1: Results**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variables | Estimate | | z-score |  |
| *Intercept* | .749 | \*\*\* | 9.471 |  |
| *Result (Lost)m* | -.086 |  | -.838 |  |
| *Result (Won)m* | .230 | \*\*\* | 3.172 |  |
| *Logged Informational MGCu,c,m* | .040 |  | .860 |  |
| *ResultLostm \* Logged Informational MGCu,c,m* | .259 | \*\*\* | 3.411 |  |
| *ResultWonm \* Logged Informational MGCu,c,m* | -.084 |  | -1.592 |  |
| *Logged Emotional MGCu,c,m* | .162 | \*\*\* | 2.944 |  |
| *ResultLostm \* Logged Emotional MGCu,c,m* | -.079 |  | -1.283 |  |
| *ResultWonm \* LoggedEmotional MGCu,c,m* | -.077 |  | -1.364 |  |
| *Unexpected Resultm* | -.083 |  | -.972 |  |
| *ResultLostm \* Unexpected m* | -.039 |  | -.310 |  |
| *ResultWonm \* Unexpected m* | .227 | \* | 1.844 |  |
| *TotalEventAttendancem* | -.121 | \*\* | -2.471 |  |
| *ResultLostm \* TotalEventAttendance m* | .128 | \*\* | 1.994 |  |
| *ResultWonm \* TotalEventAttendance m* | .163 | \*\*\* | 2.808 |  |
| *RedCardsm* | -.037 |  | -1.585 |  |
| *YellowCardsm* | -.017 |  | -.769 |  |
| *Home Game m* | -.063 |  | -1.363 |  |
| *EventFacebooku,m* | .037 |  | .557 |  |
| *Customer Sentiment u,c,m-1* | .086 | \*\*\* | 2.760 |  |
| *Other Sentiment Valence u,c,m* | .045 | \*\*\* | 3.133 |  |
| *Other Sentiment Volume u,c,m* | .060 | \*\*\* | 3.157 |  |
| *Comment length u,c,m* | -.070 | \*\*\* | -4.571 |  |
| *Comment time u,c,m* | .117 | \*\*\* | 2.812 |  |
| *ResultLost m \* Comment time u,c,m* | .078 |  | 1.491 |  |
| *ResultWon m \* Comment time u,c,m* | -.090 | \* | -1.715 |  |
| *IMRu* | .029 | \* | 1.915 |  |
| *Endogeneity Correction Logged Informational MGC m* | -.065 | \*\*\* | -2.917 |  |
| *Endogeneity Correction Logged Emotional MGC m* | -.039 | \* | -1.776 |  |
|  |  | | |  |
| *Log-Likelihood* | -5,582.1 | | |  |
| *AIC* | 11,234.1 | | |  |

Note: \* *p*<.1, \*\* *p*<.05, \*\*\* *p*<.01; coefficients are standardized. The standard errors are bootstrapped.

**Appendix W7.2: Interaction Plots for MGC**





**Appendix W8. Results of the Customer Sentiment Model with all Comments Included**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variables | Estimate | | z-score |  |
| *Intercept* | .659 | \*\*\* | 11.456 |  |
| *Result (Lost)m* | -.160 | \*\* | -2.048 |  |
| *Result (Won)m* | .369 | \*\*\* | 6.829 |  |
| *Informational MGCu,c,m* | .001 |  | .065 |  |
| *ResultLostm \* Informational MGCu,c,m* | .176 | \*\*\* | 3.692 |  |
| *ResultWonm \* Informational MGCu,c,m* | -.012 |  | -.470 |  |
| *Emotional MGCu,c,m* | .056 | \* | 1.860 |  |
| *ResultLostm \* Emotional MGCu,c,m* | -.023 |  | -.590 |  |
| *ResultWonm \* Emotional MGCu,c,m* | -.028 |  | -.891 |  |
| *Unexpected Resultm* | -.165 | \*\* | -2.503 |  |
| *ResultLostm \* Unexpected m* | .099 |  | 1.032 |  |
| *ResultWonm \* Unexpected m* | .262 | \*\*\* | 2.825 |  |
| *TotalEventAttendancem* | -.089 | \*\* | -2.340 |  |
| *ResultLostm \* TotalEventAttendance m* | .066 |  | 1.300 |  |
| *ResultWonm \* TotalEventAttendance m* | .119 | \*\*\* | 2.690 |  |
| *RedCardsm* | -.040 | \*\* | -2.362 |  |
| *YellowCardsm* | -.013 |  | -.790 |  |
| *Home Game m* | -.029 |  | -.854 |  |
| *EventFacebooku,m* | .060 |  | 1.180 |  |
| *Customer Sentiment u,c,m-1* | .070 | \*\*\* | 4.001 |  |
| *Other Sentiment Valence u,c,m* | .042 | \*\*\* | 5.330 |  |
| *Other Sentiment Volume u,c,m* | .037 | \*\*\* | 3.749 |  |
| *Comment length u,c,m* | -.081 | \*\*\* | -9.978 |  |
| *Comment time u,c,m* | .235 | \*\*\* | 11.745 |  |
| *ResultLost m \* Comment time u,c,m* | .057 | \*\* | 2.208 |  |
| *ResultWon m \* Comment time u,c,m* | -.239 | \*\*\* | -9.710 |  |
| *Endogeneity Correction Informational MGC m* | -.029 | \* | -1.688 |  |
| *Endogeneity Correction Emotional MGC m* | -.006 |  | -.457 |  |
|  |  | | |  |
| *Log-Likelihood* | -19,143.3 | | |  |
| *AIC* | 38,354.6 | | |  |
| *N (observations)\** | 37,638 | | |  |

Note: \* *p*<.1, \*\* *p*<.05, \*\*\* *p*<.01; coefficients are standardized. The standard errors are bootstrapped.

\* These observations include both customers and non-customers

**Appendix W9. Results of the Customer Sentiment Model with a One-day Timeframe**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variables | Estimate | | z-score |  |
| *Intercept* | .596 | \*\*\* | 7.328 |  |
| *Result (Lost)m* | -.309 | \*\*\* | -2.940 |  |
| *Result (Won)m* | .328 | \*\*\* | 4.305 |  |
| *Informational MGCu,c,m* | .086 | \*\* | 2.053 |  |
| *ResultLostm \* Informational MGCu,c,m* | .104 |  | .747 |  |
| *ResultWonm \* Informational MGCu,c,m* | -.117 | \*\* | -2.211 |  |
| *Emotional MGCu,c,m* | .119 | \* | 1.706 |  |
| *ResultLostm \* Emotional MGCu,c,m* | -.003 |  | -.041 |  |
| *ResultWonm \* Emotional MGCu,c,m* | -.067 |  | -.937 |  |
| *Unexpected Resultm* | -.184 | \*\* | -2.021 |  |
| *ResultLostm \* Unexpected m* | .210 | \* | 1.648 |  |
| *ResultWonm \* Unexpected m* | .395 | \*\*\* | 3.009 |  |
| *TotalEventAttendancem* | -.087 | \* | -1.767 |  |
| *ResultLostm \* TotalEventAttendance m* | .104 |  | 1.631 |  |
| *ResultWonm \* TotalEventAttendance m* | .108 | \* | 1.790 |  |
| *RedCardsm* | -.056 | \*\* | -2.392 |  |
| *YellowCardsm* | .003 |  | .126 |  |
| *Home Game m* | -.051 |  | -1.037 |  |
| *EventFacebooku,m* | .008 |  | .107 |  |
| *Customer Sentiment u,c,m-1* | .078 | \*\* | 2.186 |  |
| *Other Sentiment Valence u,c,m* | .037 | \*\* | 2.262 |  |
| *Other Sentiment Volume u,c,m* | .048 | \*\* | 2.282 |  |
| *Comment length u,c,m* | -.027 |  | -1.573 |  |
| *Comment time u,c,m* | .067 | \* | 1.764 |  |
| *ResultLost m \* Comment time u,c,m* | .064 |  | 1.298 |  |
| *ResultWon m \* Comment time u,c,m* | -.076 |  | -1.597 |  |
| *IMRu* | .038 | \*\* | 2.249 |  |
| *Endogeneity Correction Informational MGC m* | -.052 | \*\* | -2.139 |  |
| *Endogeneity Correction Emotional MGC m* | -.041 |  | -1.543 |  |
|  |  | | |  |
| *Log-Likelihood* | -4,436.5 | | |  |
| *AIC* | 8,943.0 | | |  |
| *N (observations)\** | 7,748 | | |  |

Note: \* *p*<.1, \*\* *p*<.05, \*\*\* *p*<.01; coefficients are standardized. The standard errors are bootstrapped.

\* Note that the number of observations is lower since our timeframe has shortened from 2 days to 1 day. Hence, posts by the team (and the comments by customers) during day two are not taken into account.

**Appendix W10. CLV Model Specifications**

**Appendix W10.1: Data and Model Set-up**

***Data***

Next to the data collected for the customer sentiment model, we aim to include page likes on social media (share of interest variable, see below). In order to achieve this, we built an application on Facebook. The team hosted and advertised the app on its Facebook page. To encourage app usage, we offered a chance to win a prize (an autographed shirt). Once users clicked on a link, an authorization box required them to give permission for gathering data; they were told what would be collected. Once opened, the app presented three team-related questions and a tie-breaker to identify a winner. The app collected data on users (age, gender, location, and email address) and SM (page likes, comments, and post likes). See our modeling discussion for further details.

Using the application, we were able to collect data for 5,783 customers. These customers have used the app and commented on the Facebook page in our specific research window (2011-2014). Thus, for our CLV model, we also use 4 years of information. We address sample selection issues related to the app usage below.

***Model***

We model CLV following the choice-then-quantity approach (Kumar, Venkatesan, Bohling and Beckmann 2008). This model follows the always-a-share approach to measuring CLV, assuming customers never terminate their relationship but may have dormancy periods. This is also supported by the data, as 25% of the customers who stop buying a season ticket restart at some point in the future without specific targeting efforts by the company. A lost-for-good approach would thus result in a serious underestimation of CLV (Gupta et al. 2006). A customer’s CLV is operationalized as the net present value of his/her future cash flows, computed over three years. Following (Kumar et al. 2008), CLV is modeled as

|  |  |
| --- | --- |
|  | *( W10.1.1)* |

where

CLV *i* = CLV for customer *i*,

p(*Purchaseit*) = predicted probability of purchase for customer *i* in year *t*

= predicted purchase amount of customer *i* in period *t*[[5]](#footnote-6)

= average cost for a single communication (email, estimated to be € 0.89 by the soccer team)

= predicted marketing contacts (e-mails) for customer *i* in year *t*

*t* = year index

*T* = end of the observation phase, and

*R* = yearly discount rate (0.15 as is common in CLV studies)

The CLV formula in Equation W10.1.1 consists of all revenue a buyer accrues minus marketing costs. While the number of marketing contacts is commonly considered endogenous (e.g., Kumar et al. 2008), the team contacts (via undirected e-mails) all customers. For verification, we form three spending groups—low (student rate), average, and high (VIP); and calculate the mean number of e-mails sent. Results reveal no differences across groups, confirmed by an ANOVA (F = .4881, *p* = .4849), showing there is no need to account for endogeneity. We also consider possible systemic differences in contacts over time. For the last year in our data, the number of e-mails is higher than in previous years due to a change in the team’s marketing agency. However, there are no differences across groups in that year. Thus, we model predicted marketing contacts in a constant way because there are no specific communications drivers (MTi0 = MTi1 = MTi2 =…); but we include interaction during the last year (2014) and contact volume to account for time-specific changes. Finally, we can rule out the possibility of intra-year differences (e.g. strategically timing e-mails) since the e-mails are sent on average every two weeks, coinciding with the home games and providing information on these games.

The CLV formula requires predicting two other concepts: (1) purchase probability and (2) purchase amount of customer *i* in year *t*. Purchase amount is only observed if buyers purchase. In the purchase equation, we assume customer *i* will buy only when the latent utility exceeds zero. However, we do not observe this latent utility; instead, we observe only the buy vs. no-buy decision. We map the latent utility to the decision using a binary probit choice model:

|  |  |
| --- | --- |
|  | *( W10.1.2 )* |

Then, we model the latent utility as a linear function of the predictor variables:

|  |  |
| --- | --- |
|  | *( W10.1.3 )* |

where is a vector of customer specific intercepts, is a vector of coefficients, is a vector containing predictor variables listed in Equation W.10.1.6, and captures the error term. Similarly, we assume the latent variable to represent the amount of purchases of customer *i* in period *t*:

|  |  |
| --- | --- |
|  | *( W10.1.4 )* |

where is again a vector of customer specific intercepts, is a vector of coefficients, is a vector containing predictor variables for the purchase amount equation (W.10.1.7), and captures the error term. The latent amount is observed when a customer purchases:

|  |  |
| --- | --- |
| *if* , otherwise unobserved. | *( W10.1.5 )* |

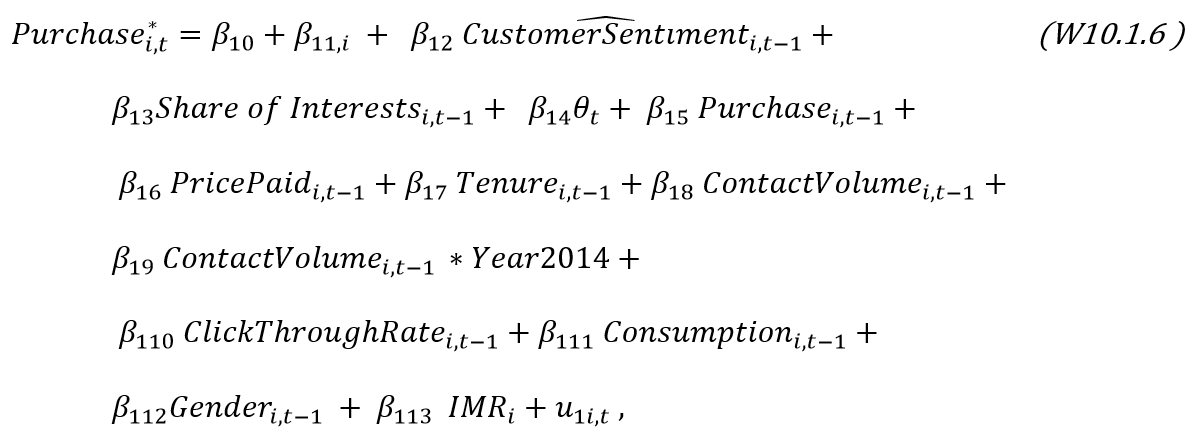
Our dataset consists of panel data, with purchases over time specified in the purchase incidence and amount equations. Hence, we cannot apply the simple selection model; instead, we use the random-effects variant (Bruce, Desai, and Staelin 2005; Verbeek and Nijman 1992) in Limdep 11. Parameters and in equations W.10.1.3 and W.10.1.4 denote random effects (not simple intercepts), assumed to be bivariate normally distributed with zero means, s.d. and and correlation θ. We specify the random-effects variant because selectivity comes from correlation of the errors and and random effects and (Greene 2016a). We jointly fit equations via maximum simulated likelihood in Limdep, implying no IMR variable for this selection bias.

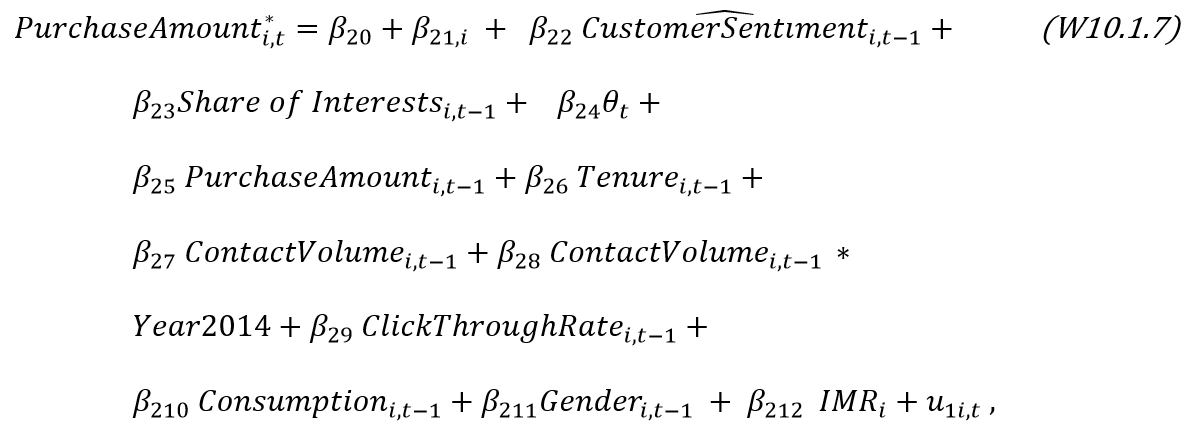
***Variables***

Research has established the importance of CLV due to its impact on firm value (Gupta, Lehmann, and Stuart 2004) as well as its significance for developing a competitive advantage (Brodie, Ilic, Juric and Hollebeek 2013), driving sales growth (Voyles 2007), and efficiently allocating firm resources (Kumar et al. 2008). Pansari and Kumar (2017) maintain that CLV is the relevant metric for assessing customers’ direct engagement with firms, defined as the value of a customer’s future contributions to the firm. We measure CLV based on both purchase incidence and purchase amount.

The link between customer sentiment related to events and behavioral outcomes such as purchases is established in the literature. Consistent with the service-profit chain (Anderson and Mittal 2000), which has been applied to sports contexts (Wetzel et al. 2018), satisfaction as a consequence of positive event outcomes influences purchase behavior (Bolton 1998). In addition, satisfaction is a key requirement for customer engagement and influences buyers’ purchase behavior and, hence, their CLV (Pansari and Kumar 2017). Thus, we anticipate customer sentiment in SM, as a result of customer interactions, to be related to CLV. Note that we do not imply a causal relationship, but rather evaluate customer sentiment as a leading indicator, similar to previous work exploring social media activity as a leading indicator of later customer behavior (e.g., Schweidel and Moe 2014; Srinivasan, Venhuele and Pauwels 2010). We further expect this relationship to hold beyond the impact of buyers’ previous activities with the firm, which are captured with control variables.

We include one control variable based on social media, which is the team’s *share of interests*, operationalized as the percentage of all Facebook page likes (at time *t-1*) by the users that are related to the team (e.g., liking the official fan page of the team, liking players’ pages, liking unofficial fan pages discussing the team). Previous work finds similar notions such as share of wallet to influence CLV (Reinartz, Thomas, and Kumar 2005) as a control variable, which we apply to the SM domain. This expectation is consistent with arguments proposing that firms owning a greater share of customers’ wallets enjoy stronger relationships based on longer relationships and enhanced abilities to learn about buyer needs via more communication (Anderson and Narus 2003).

Finally, we add control variables related to demographics and several aspects of customer-team interactions, which are based on previous CLV literature (e.g., Kumar et al. 2008; Reinartz, Thomas, and Kumar 2005), literature on retention or attrition (e.g., Ascarza, Netzer, and Hardie 2018; Benoit and Van den Poel 2012; Nitzan and Libai 2011) and literature investigating the purchase amount (e.g., Goh, Heng, and Lin 2013; Kumar et al. 2016). The buying equation includes a lagged purchase indicator (*Purchaset-1*), lagged mean purchase amount (*Paid Price*), buyer tenure (*Tenure*), gender (*Gender*), number of e-mails sent to the customer (*Contact Volume*), and email click-rate (*Click-Through Rate*). The purchase amount equation includes the lagged purchase amount (*Purchase Amountt-1*). Note that the only variables that are different in the purchase incidence versus purchase incidence equations are *Purchaset-1* and *Paid Price* versus *Purchase Amountt-1*. The reason is twofold. First, we believe that these variables are most informative for the respective dependent variable. Second, not having the exact same independent variables allows the model to identify all variables. Finally, we also add a variable to capture the ‘usage’ or ‘consumption’ of the season ticket, by including the lagged percentage of home matches attended (*Consumption*) to each of these equations. This is similar to the notions of usage in the cellular phone industry (e.g., Nitzan and Libai 2011). The final equations take the following form: 



The second term in each equation represents the customer-specific intercept; and the variable represents a vector of year-specific intercepts, accounting for factors that might vary by year. Thus, the intercept is both customer-specific and time varying. The inverse Mills ratio in both equations represents the inverse Mills ratio calculated from the selection equation (see further). No second inverse Mills ratio factor is in the purchase amount equation; the second inverse Mills ratio factor would come from the selection based on purchase incidence, but these equations are jointly estimated by maximum simulated likelihood. This estimation does not use a two-step method and hence does not create or use an inverse mills ratio variable. Section 2 of this web appendix details the estimation procedure. Section 3 lists variables in the CLV model. Section 4 provides the CLV model’s descriptive measures, distributions of both purchase incidence and purchase amount, and correlation matrices. The last sections contain results and detailed descriptions of these results.

*Self-selection issue.* Similar to the customer sentiment regression, our data may suffer from sample selection bias because customers included in the CLV analysis self-selected into the study by allowing the app to extract their additional Facebook information. These individuals may not be representative of the population because there may be unobserved factors influencing both the decision to use the app and their buying behavior. This self-selection potentially leads to an endogeneity issue due to omitted variables bias (Wies and Moorman 2015), alleviated by implementing a binary probit choice model as a Heckman selection model (Heckman 1979)[[6]](#footnote-7).

It may not be surprising that the large majority of people who used the app also commented on the Facebook page (95% of the app users who were matched with the internal database also commented on the Facebook page). We proceed with the 5,871 customers that also commented. The selection regression used here is similar to the selection regression used for the customer sentiment regression, with the only difference being the focus on season ticket holders only (instead of both season and loose ticket holders in the customer sentiment equation). While we keep most of the explanatory variables, we change the online purchase variable with privacy related variables. The reason is that it is more likely that online purchases will influence purchase incidence and amount. Privacy related information on the other hand is unlikely to influence purchase behavior, while we expect that those more concerned about privacy may not be active in SM and hence will be less likely to use the application (Kumar et al. 2016). We use disclosure of a telephone number and identity card number as a proxy for privacy concern (Goh, Heng and Lin, 2013). We again stress that, in order to satisfy the exclusion restrictions, the selection equation needs at least one significant independent variable that will not affect the final dependent variable (Puhani, 2000), which is achieved by including demographics and privacy-related information. Note that our exclusion restrictions are fulfilled, and that the CLV model estimates can be correctly identified, even if one of the variables (customer tenure) is also included in the selection regression. Since the procedure is exactly similar to the one outlined in the main body of the paper, we do not repeat the steps here. The CLV model can be seen as the second step of the selection model, which depends on the selection equation. By including the inverse mills ratio in the CLV model equations as an explanatory variable, we correct for potential endogeneity issues resulting from self-selection. If the inverse Mills ratio coefficient is significant, self-selection is indeed an issue.

Please note that we have two different selection issues in the complete CLV model. The first selection issue evaluates whether or not customers are using the team’s application or not. The second selection arises from the fact that spending (purchase amount) is only observed when purchases occur. In order to solve the first selection issue, we use a two-step Heckman model and include the inverse Mills ratio based on the selection equation for app usage in the CLV model (purchase incidence and purchase amount equations). In order to solve the second selection issue, instead of a two-step Heckman model, we use a model that uses maximum likelihood optimization. The reasons for this difference in approach between the two selections are:

1) The first selection issue determines whether a customer has used the application. Thus, we only have one selection observation per panel group (=customer) in our CLV model, and the specific panel data structure can be neglected here. For this reason, we use the more straightforward two-step Heckman method.

2) The second issue is a complicated selection since we are not only dealing with the selection issue but also with panel data in our selection. That is, the selection takes place within each of our panels (e.g., customer X can buy in 2011, but not in 2012). While previous research (e.g., Goh, Heng and Lin, 2013; Wies and Moorman, 2015) has neglected this important complication and used the simpler (but in this case, incorrect) two-step model, we opted to use a more accurate model that correctly considers the selection.

3) Combining all equations (application usage regression, purchase incidence and purchase amount equations) and estimating one model with maximum likelihood optimization would greatly increase the complexity of the model. These types of models do not exist yet.

**Appendix W10.2: Detailed Model Description**

The CLV model estimation is based on the Limdep implementation of the RE sample selection model (Greene, 2016a). Starting with the purchase incidence and amount equations:

|  |  |
| --- | --- |
|  | *(W.1)* |
|  | *(W.2)* |
|  | *(W.3)* |

where and are vectors of customer specific intercepts, and are vectors of coefficients, and are vectors containing the predictor variables and and capture the error terms in the buying and purchase amount respectively ( ~ N[0,1] and ~ N[0,]). Let ρ = cor(,), then the contribution of group *i* to the log likelihood can be described as:

|  |  |
| --- | --- |
|  | *(W.4)* |

However, and are unobserved. We therefore obtain the unconditional log likelihood by integrating out the random effects:

|  |  |
| --- | --- |
| Let , | *(W.5)* |
| Then, | *(W.6)* |

In order to solve this, Monte Carlo simulation is used and the integral is approximated by

|  |  |
| --- | --- |
|  | *(W.7)* |

where , are R random draws from the joint distribution of and . The approximation improves with increasing R. The simulation allows for two parameters to be set: the method of random draws and the number of draws. Based on the recommendations in Greene (2016b) we use 1000 Halton draws (for a more in depth discussion of Halton draws, see Greene (2016b) and Train (1999)).

Then, the total log likelihood can be described as:

|  |  |
| --- | --- |
|  | *(W.8)* |

This likelihood function is then maximized by solving the likelihood equations:

|  |  |
| --- | --- |
|  | *(W.9)* |

where Θ refers to the vector of parameters in the model. These derivatives must be approximated as well. Please see Greene (2016b) for a detailed description of the process.

**Appendix W10.3: Variables in the CLV Model**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Description** |  | |  | |
| ***CLV model*** |  | |  | |  | |
| *Dependent Variables* | *Measures of CLV* | | *Purchase incidence* | | *Purchase Amount* | |
| *Purchaset* | Dummy indicating whether a season ticket was bought at time *t* | | X | |  | |
| *PurchaseAmountt* | Amount spent on season tickets at time *t (€)* | |  | | X | |
| *Variables of interest* | *Focal Measures* | |  | |  | |
| *t-1* | Predicted customer sentiment during period *t*-1 | | X | | X | |
| *Control variables* | *Control variables for the CLV models* | |  | |  | |
| *θt* | Dummy variables indicating the year *t* | | X | | X | |
| *Share of Interests t-1* | Percentage of page likes on Facebook during related to the team *t*-1 | | X | | X | |
| *Tenure t-1* | Length of relationship of the focal customer with the company at time *t*-1 (in years) | | X | | X | |
| *Purchaset-1* | Dummy indicating whether a season ticket was bought at time *t*-1 | | X | |  | |
| *PurchaseAmountt-1* | Amount spent on season tickets at time *t*-1 (€) | |  | | X | |
| *PricePaid t-1* | Average amount spent on season tickets by the focal customer at time *t-*1(€) | | X | |  | |
| *ContactVolume t-1* | Number of email messages sent by the company to the focal customer during period *t*-1 (logarithm) | | X | | X | |
| *ContactVolume t-1 \* Year2014* | Interaction effect of contact volume and the year 2014 (to account for change in marketing agency) | | X | | X | |
| *Click-through Rate t-1* | Click through rate of the customer on emails received by the company during period *t-*1 (percentage) | | X | | X | |
| *Consumptiont-1* | The percentage of total (home) matches attended by the customer during period *t*-1 | | X | | X | |
| *Gender* | Gender of the focal customer (dummy) | | X | | X | |

**Appendix W10.4: Descriptives of Variables Included in the CLV Model**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Purchase incidence** | | | **Purchase amount** | | | |
| *Variables* | *M* | *SD* | *RANGE* | *M* | *SD* | *RANGE* | |
| *Purchase incidence* | .65 | .48 | [0,1] |  |  |  |
| *Purchase amount* |  |  |  | 256.9 | 124.96 | [.01,3205] |
|  | .61 | .11 | [0.35,1] | .61 | .11 | [.35,.93] |
| *Share of interests* | .07 | .14 | [0,2] | .08 | .14 | [,1.75] |
| *Purchaset-1* | .64 | .48 | [0,1] |  |  |  |
| *PurchaseAmountt-1* |  |  |  | 200.28 | 142.01 | [0,3120.02] |
| *Price Paid* | 144.42 | 130.22 | [0,814] |  |  |  |
| *Tenure* | 4.92 | 3.10 | [0,10] | 5.19 | 2.99 | [0,10] |
| *Contact Volume\** | .43 | .62 | [0,1.65] | .46 | .64 | [0,1.65] |
| *Click-Through Rate* | .02 | .07 | [0,1] | .02 | .07 | [0,1] |
| *Consumption* | .26 | .33 | [0,1] | .35 | .35 | [0,1] |
| *Gender* | .91 | .29 | [0,1] | .90 | .30 | [0,1] |

\*logarithm

**Correlation Table Purchase Incidence Equation**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Variable* | | *1* | *2* | *3* | *4* | *5* | *6* | *7* |
| *1* |  | 1 |  |  |  |  |  |  |
| *2* | *Share of interests* | .101 | 1 |  |  |  |  |  |
| *3* | *Price Paid* | .075 | .087 | 1 |  |  |  |  |
| *4* | *Tenure* | .023 | .026 | .269 | 1 |  |  |  |
| *5* | *Contact Volume* | -.008 | .010 | .151 | .185 | 1 |  |  |
| *6* | *Click-Through Rate* | -.028 | .030 | .099 | .076 | .369 | 1 |  |
| *7* | *Consumption* | -.021 | .085 | .379 | .240 | .383 | .207 | 1 |

Note: Only continuous variables are reported; correlations with absolute values above 0.0129 are significant at a 5% significance level (N = 23,132)

**Correlation Table Purchase Amount Equation**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Variable* | | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* |
| *1* | *Purchase amount* | 1 |  |  |  |  |  |  |  |
| *2* |  | .004 | 1 |  |  |  |  |  |  |
| *3* | *Share of interests* | .013 | .097 | 1 |  |  |  |  |  |
| *4* | *PurchaseAmountt-1* | .609 | .068 | .035 | 1 |  |  |  |  |
| *5* | *Tenure* | .204 | .004 | .004 | .287 | 1 |  |  |  |
| *6* | *Contact Volume* | .194 | -.007 | .003 | .152 | .163 | 1 |  |  |
| *7* | *Click-Through Rate* | .115 | -.038 | .022 | .071 | .068 | .391 | 1 |  |
| *8* | *Consumption* | .291 | -.048 | .050 | .376 | .267 | .385 | .214 | 1 |

Note: Only continuous variables are reported; correlations with absolute values above 0.016 are significant at a 5% significance level (N = 15,061)

**Appendix W10.5: CLV Model Results**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Purchase incidence** | | | | **Purchase amount** | | | | | | | |
| *Variables* | *Estimate* |  | *z-score* | | *Estimate* | | |  | | | *z-score* | |
| *Intercept* | -.28 | \*\*\* | -6.01 | | 164.90 | | \*\*\* | | | 58.63 | | |
| *Year: 2012* | -.07 | \*\* | -2.37 | | -6.13 | | \*\* | | | -2.35 | | |
| *Year: 2013* | -.02 |  | -.53 | | 70.59 | | \*\*\* | | | 21.81 | | |
| *Year: 2014* | -.22 | \*\*\* | -6.43 | | 27.68 | | \*\*\* | | | 9.69 | | |
|  | .03 | \*\* | 2.38 | | .45 | |  | | | .50 | | |
| *Share of Interests* | .06 | \*\*\* | 6.45 | | 1.80 | | \*\* | | | 2.44 | | |
| *Purchaset-1* | 1.09 | \*\*\* | 30.74 | |  | |  | | |  | | |
| *PurchaseAmountt-1* |  |  |  | | 79.48 | | \*\*\* | | | 309.26 | | |
| *Price paid* | .09 | \*\*\* | 5.11 | |  | |  | | |  | | |
| *Tenure* | .01 |  | 1.32 | | 5.14 | | \*\*\* | | | 5.92 | | |
| *Contact Volume* | -.05 | \*\*\* | -3.25 | | -1.34 | |  | | | -1.04 | | |
| *Year2014\*Contact Volume* | .07 | \*\*\* | | 3.20 | 4.01 | | \*\* | | 2.29 | | | |
| *Click-through Rate* | .05 | \*\*\* | 5.30 | | 3.82 | | \*\*\* | | | 4.54 | | |
| *Consumption* | .18 | \*\*\* | 11.43 | | .93 | |  | | | .83 | | |
| *Gender* | .03 |  | 1.02 | | 8.68 | | \*\*\* | | | 3.87 | | |
| *IMR* | .02 | \*\* | 2.37 | | 13.14 | | \*\*\* | | | 20.20 | | |
|  |  |  |  | | |  | | | | | |  |
| *σ* | .01 .06 | | | | | | | | | | | |
| *ρ* | .87 \*\*\* 281.21 | | | | | | | | | | | |
| *AIC* | 187,461.3 | | | | | | | | | | | |
| *Log-Likelihood* | -93,697.66 | | | | | | | | | | | |

Note: \* *p*<.1, \*\* *p*<.05, \*\*\* *p*<.01; coefficients are standardized; subscripts are not included for clarity, except for the lagged dependent variables

**Appendix W10.6: Discussion of CLV Model Results**

The CLV model’s results are shown in Web Appendix W10.5. As the models are jointly estimated, only one log-likelihood and AIC value are available per joint buying and purchase amount model. All but one time-varying intercept are significant, indicating that time-varying effects may be necessary to absorb season-specific shocks.

Predicted customer sentiment has the expected (positive) direction in both equations ( = .03 and = .45). However, the effect is only significant in the purchase incidence equation. Thus, we conclude that the expressed sentiment on social media can be seen as a leading indicator of the probability of purchasing a ticket for the next season, but that it is not linked to the amount paid for the season ticket (that is, people will not upgrade their season ticket with higher sentiment). We emphasize that this result holds while controlling for typical strong indicators of CLV and for online customer share of interest. We note, however, that the typical transactional indicators are more strongly related to CLV than our customer sentiment measure.

We include one other SM control variable, share of interests. We see that a higher percentage of likes related to the company is linked to both a higher probability of purchase and a higher purchase amount ( = .06, *p* < .01; and = 1.80, *p* < .05). Thus, next to the expressed sentiment on social media, social identification with the team on social media can also be seen as an indicator of CLV. Note that our measure is broader than simply liking the official team page; it covers all Facebook fan pages related to the team. As a robustness check, we estimated the model with a dummy variable (indicating whether or not a customer has liked the official page) instead of share of interest. In this case, the ‘liking’ variable was not significant.

Most of the other control variables are significant for both purchase incidence and purchase amount equations. We first focus on purchase incidence. Prior research reports positive impacts of previous purchase behavior and price on purchase probability, both of which are confirmed (𝛽15  = 1.09 and 𝛽16  = .09, respectively; both *p* < .01). Tenure (𝛽17 however, is not significant. Prior research also suggests that marketing communications positively relate to purchase probability, which is only partially confirmed: contact volume (𝛽18 = -.05, *p* < .01) negatively relates to purchase incidence, but with a significant, positive interaction effect in 2014 (𝛽19 = .07, *p* < .01). Click-through rate is positive and significant (𝛽110 = .05, *p* < .01). We confirm that the percentage of matches attended positively relates to purchase incidence (𝛽111 = .18, *p* < .01). Finally, we note that gender is not significant.

For the purchase amount equation, all controls are significant and have the expected sign except for contact volume, which is not significant but has a significant positive interaction effect in 2014 (𝛽28 = 4.01, *p* < .05). The amount spent on season tickets last year is most important.

Furthermore, the inverse mills ratio is significant for both equations, indicating that self-selection is an issue. A higher inverse mills ratio value indicates a lower probability to use the app. Given the positive sign of the inverse mills ratio parameter coefficients ( = .02 and = 13.14), we conclude that the lower the probability to use the app, the higher the propensity to purchase and the higher the average purchase amount. This finding is not surprising since a large percentage of the sample are younger and have lower season ticket expenditures (average season ticket price for sample customers is €147 vs. €155 for out-of-sample customers). The logical interpretation of the inverse mills ratio coefficients adds face validity to the results.

Finally, σ and ρ represent the standard error for the purchase amount equation and the correlation between residuals of the purchase amount and incidence equations. We use the parameters in estimating the panel selection model for the equations. The high correlation (ρ) suggests the need to use a panel selection specification to model purchase amount and incidence.

**Web Appendix References**

Anderson, Eugene W. and Vikas Mittal (2000), “Strengthening the Satisfaction-Profit Chain,” *Journal of Service Research*, 3 (2), 107–20.

Anderson, James C and James A. Narus (2003), “Selectively pursuing more of your customer’s business,” *MIT Sloan Management Review*, 44 (3), 4436.

Ascarza, Eva, Oded Netzer, and Bruce G. S. Hardie (2018), “Some Customers Would Rather Leave Without Saying Goodbye,” *Marketing Science*, 37 (1), 54-77

Bartling, Björn, Leif Brandes, and Daniel Schunk (2015), “Expectations as Reference Points: Field Evidence from Professional Soccer,” *Management Science*, 61 (11), 2646–61.

Benoit, Dries F. and Dirk Van den Poel (2012), “Improving Customer Retention in Financial Services using Kinship Network Information,” *Expert Systems with Applications*, 39 (13), 11435–42.

Beukeboom, Camiel J., Peter Kerkhof, and Metten de Vries (2015), “Does a Virtual Like Cause Actual Liking? How Following a Brand’s Facebook Updates Enhances Brand Evaluations and Purchase Intention,” *Journal of Interactive Marketing*, 32, 26–36.

Bolton, Ruth N. (1998), “A Dynamic Model of the Duration of the Customer’s Relationship with a Continuous Service Provider: The Role of Satisfaction,” *Marketing Science*, 17 (1), 45–65.

Brodie, Roderick J., Ana Ilic, Biljana Juric, and Linda Hollebeek (2013), “Consumer Engagement in a Virtual Brand Community: An Exploratory Analysis,” *Journal of Business Research*, 66 (1), 105–14.

Bruce, Norris, Preyas S. Desai, and Richard Staelin (2005), “The Better They Are, the More They Give: Trade Promotions of Consumer Durables,” *Journal of Marketing Research*, 42 (1), 54–66.

Card, David and Gordon B. Dahl (2011), “Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior,” *The Quarterly Journal of Economics*, 126 (1), 103–43.

Dolton, Peter and George Mackerron (2018), “Is Football a Matter of Life and Death - Or Is It More Important Than That?,” *NIESR Discussion Paper No. 493*.

Farhadloo, Mohsen, Raymond A. Patterson, and Erik Rolland (2016), “Modeling Customer Satisfaction from Unstructured Data using a Bayesian approach,” *Decision Support Systems*, 90, 1–11.

Gijsenberg, Maarten J., Harald J. Van Heerde, and Peter C. Verhoef (2015), “Losses Loom Longer than Gains: Modeling the Impact of Service Crises on Perceived Service Quality over Time,” *Journal of Marketing Research*, 52 (5), 642–56.

Goh, Khim-Yong, Cheng-Suang Heng, and Zhijie Lin (2013), “Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User- and Marketer-Generated Content,” *Information Systems Research*, 24 (1), 88–107.

Greene, William .H., (2016a), “Sample Selection Models for Panel Data,” in*: Econometric Modeling Guide Limdep 11.* Econometric Software, Inc.

-------- (2016b). “Random Parameter Models”, in: *Limdep 11 Reference Guide*. Econometric Software, Inc.

Gupta, Sunil, Dominique Hanssens, Bruce Hardie, Wiliam Kahn, V. Kumar, Nathaniel Lin, Nalini Ravishanker, and S. Sriram (2006), “Modeling Customer Lifetime Value,” *Journal of Service Research*, 9 (2), 139–55.

Gupta, Sunil, Donald R. Lehmann, and Jennifer Ames Stuart (2004), “Valuing Customers,” *Journal of Marketing Research*, 41 (1), 7–18.

Hart, Andrew, Finola Kerrigan, and Dirk vom Lehn (2016), “Experiencing Film: Subjective Personal Introspection and Popular Film Consumption,” *International Journal of Research in Marketing*, 33 (2), 375–91.

Heckman, James J. (1979), “Sample Selection Bias as a Specification Error,” *Econometrica*, 47 (1), 153–61.

Kumar, V., Rajkumar Venkatesan, Tim Bohling, and Denise Beckmann (2008), “The Power of CLV: Managing Customer Lifetime Value at IBM,” *Marketing Science*, 27 (4), 585–99.

Kumar, Ashish, Ram Bezawada, Rishika Rishika, Ramkumar Janakiraman, and P.K. Kannan (2016), “From Social to Sale: The Effects of Firm-Generated Content in Social Media on Customer Behavior,” *Journal of Marketing*, 80 (1), 7–25.

Nam, Sungjoon, Puneet Manchanda, and Pradeep K. Chintagunta (2010), “The Effect of Signal Quality and Contiguous Word of Mouth on Customer Acquisition for a Video-on-Demand Service,” *Marketing Science*, 29 (4), 690–700.

Nitzan, Irit and Barak Libai (2011), “Social Effects on Customer Retention,” *Journal of Marketing*, 75 (6), 24–38.

Pansari, Anita and V. Kumar (2017), “Customer Engagement: The Construct, Antecedents, and Consequences,” *Journal of the Academy of Marketing Science*, 45 (3), 294–311.

Puhani, Patrick (2000), “The Heckman Correction for Sample Selection and Its Critique,” *Journal of Economic Surveys,* 14 (1), 53–68.

Reinartz, Werner, Jacquelyn S. Thomas, and V. Kumar (2005), “Balancing Acquisition and Retention Resources to Maximize Customer Profitability,” *Journal of Marketing*, 69 (1), 63–79.

Schweidel, David A. and Wendy W. Moe (2014), “Listening In on Social Media: A Joint Model of Sentiment and Venue Format Choice,” *Journal of Marketing Research*, 51 (4), 387–402.

Srinivasan, Shuba, Marc Vanhuele, Koen Pauwels, (2010), “Mind-Set Metrics in Market Response Models: An Integrative Approach,” *Journal of Marketing Research*, 47 (4), 672–84.

Train, K. (1999), “Halton Sequences for Mixed Logit,” Manuscript, Department of Economics, University of California, Berkeley.

Verbeek, Marno and Theo Nijman (1992), “Testing for Selectivity Bias in Panel Data Models,” *International Economic Review*, 33 (3), 681–703.

Voyles, Bennett (2007), “Beyond loyalty: Meeting the Challenge of Customer Engagement,” *Economist, Intelligence Unit*.

Wetzel, Hauke A., Stefan Hattula, Maik Hammerschmidt and Harald J. van Heerde (2018), “Building and Leveraging Sports Brands: Evidence from 50 Years of German Professional Soccer,” *Journal of the Academy of Marketing Science*, 46 (4), 591-611.

Wies, Simone and Christine Moorman (2015), “Going Public: How Stock Market Listing Changes Firm Innovation Behavior,” *Journal of Marketing Research*, 52 (5), 694–709.

1. These odds are also called 1X2 odds. The odds were gathered using the website <https://www.oddsportal.com/>, which captures odds from different bookmakers and takes the average which means we do not rely on one specific bookmaker. The pre-play odds are closed just before the start of the match, thus taking into account all information that is also available to customers. [↑](#footnote-ref-2)
2. The sum of the inverse odds results in 1.054. Thus, we have to adjust the inverse odd of a win (1/1.40 = .7142) by dividing it by the total sum of the inverse odds. [↑](#footnote-ref-3)
3. This is also a logical finding, since there are more chances for wins and losses to occur compared to draws (e.g., a draw can only be 0-0, 1-1, 2-2 etc, while a win can be 1-0, 2-0, 3-0 etc but also 2-1, 3-1, 3-2, etc). [↑](#footnote-ref-4)
4. In the sample, we had 66 draws; 1/3 was classified as expected, and 2/3 as unexpected. As an alternative, we could look to the cases where the calculated probability of a draw was higher than either the probability of a win or the probability of a loss. This resulted in only 1 different case and did not change the results of our analyses. [↑](#footnote-ref-5)
5. We use purchase amount as opposed to contribution margin because the team’s costs are fixed, unknown, and do not vary among customers. [↑](#footnote-ref-6)
6. We do not use purchase amount as an independent variable in the selection equation because of endogeneity concerns (Wies and Moorman 2015) (i.e., purchase amount is also used as the dependent variable in CLV equation). [↑](#footnote-ref-7)