Schmollers Jahrbuch 136 (2016), 199 – 226
Duncker & Humblot, Berlin

Volunteering, Match Quality, and Internet Use

By Eike Emrich* and Christian Pierdzioch**

Abstract

We use boosted regression trees to study the interplay between match quality and Internet use of volunteers. Match quality reflects the congruence of volunteers' motives for doing volunteer work and their utility experiences. Using data from an online survey questionnaire of volunteers working for the German Red Cross, we find a positive correlation between match quality and social media use and, to a lesser extent, the intensity of volunteering-related Internet use. We study the relative importance of Internet use and other control variables for match quality, the partial dependence of match quality on Internet use and the control variables, and the interaction of Internet use with the control variables.

JEL Codes: H41, J22, L31

1. Introduction

In contrast to a widely held reservation that mass media usage (Putnam 1995) in general and the adoption of the Internet in particular destroy social capital, results of a significant and growing body of research show that Internet use has a neutral or even a positive effect on the social integration of users and their civic and political participation (Hampton and Wellman 2003; Franzen 2003; Bauernschuster et al. 2014; among others). With regard to volunteering, Emrich et al. (2014) show that volunteer labor supply exhibits a positive correlation with the intensity of Internet use, and Emrich and Pierdzioch (2016) find that the intensity of volunteering-related use of the Internet exhibits a positive correlation with volunteers' commitment and their satisfaction with their volunteer work. In our empirical analysis, we go beyond earlier research

* Department of Sports Science, Saarland University, Campus Buildung B8.2, 66123 Saarbrücken, Germany. The author can be reached at e.emrich@mx.uni-saarland.de.

** Corresponding author. Department of Economics, Helmut Schmidt University, Holstenhofweg 85, PO Box 700822, 22008 Hamburg, Germany. The author can be reached at c.pierdzioch@hsu-hamburg.de.

We thank two anonymous reviewers for helpful comments. The usual disclaimer applies.

in that we study one specific channel through which Internet use may be linked to volunteers' commitment: the matching of volunteers' motives and their utility experiences.

As with employer-employee relations in for-profit organisations studied in the large and significant labor-economics literature on search-and-matching models, volunteering in non-profit organisations can be interpreted as a process of search (Schiff 1980). This process has at least two dimensions. First, a volunteer *in spe* needs to find an organisation that is a good match for his or her humanitarian or ideological values. Second, once a volunteer has found such an organisation, a volunteer needs to learn about how this organisation works and whether he or she can get along with the staff and other volunteers working for this organisation. Importantly, a volunteer needs to find a specific position that makes it possible to match utility experiences from doing volunteer work with his or her motives for doing volunteer work. A mismatch of motives and utility experiences can be expected to lead to dissatisfaction, reduced work effort, weak organisational bonds, and eventually a high turnover rate. This research sheds light on the second dimension of the search process by studying empirically whether specific forms of Internet use and the intensity of Internet use correlate with match quality in the German Red Cross (GRC), a large non-profit organisation.

In order to measure match quality, we combine data on motives for doing volunteer work with data on utility experiences – both collected by means of an online survey questionnaire – to form an index of match quality. We measure motives and utility experiences along three dimensions that cover the three broad types of models commonly studied in economic research on volunteering: the public-goods model (Roberts 1984; Bergstrom et al. 1986; Duncan 1999), the private-consumption model (Andreoni 1989 and 1990; Harbough 1998), and the human-capital model (Menchink and Weisbrod 1987). Technically, we construct our index of match quality along the lines suggested by Stukas et al. (2009). They construct an index of match quality by combining data on volunteer motives with data on environmental affordances to meet those motivations, where volunteer motives are measured using a functional approach developed in an influential contribution by Clary et al. (1998). While the match quality index that Stukas et al. (2009) propose reflects the research tradition in the psychology literature, our index of match quality brings together data on volunteer motives and volunteer utility experiences and thus draws on the terminology of economics.

While it is an interesting question whether the Internet helps to balance volunteer motives and volunteer utility experiences by improving search-and-matching efficiency, Internet use certainly is not the only variable influencing match quality. While earlier research in labour economics on the determinants of job satisfaction (e.g. Green 2010; Gielen 2013) and in sociology on volunteer satisfaction (e.g. Cnaan and Cascio 1999; Boezeman and Ellemers 2008;

Garner and Garner 2009; Hustinx 2010; for a survey of the literature, see Wilson 2012) provides some guidance as to which variables may influence match quaility, in principle a potentially large group of predictor variables may influence match quality to different extents. Some elements in this group should reflect how volunteer work is organised. For example, match quality could be linked to whether a volunteer receives support from the GRC, whether a volunteer can participate in communication and decision processes, whether training programmes for volunteers are available, and whether a volunteer has a flexible timetable for doing his or her work. Other elements in the group of predictor variables should reflect that match quality could be linked to various socioeconomic variables like a volunteer's social integration into the GRC group, a volunteer's general humanitarian values, and last but not least the overall importance of the volunteer work at the GRC for a volunteer. Match quality could be linked to all these predictor variables in a complex and potentially nonlinear way, and it is also conceivable that the various predictor variables interact with one another. For example, Internet use may strengthen the effect of training programmes on match quality because it brings these training programmes to a volunteer's awareness, helps to establish efficient e-learning techniques, and opens up the possibility that a volunteer gives feedback during a training programme.

We apply a flexible machine learning algorithm known as boosted regression trees (Friedman 2001 and 2002) to study how match quality is linked to a broad range of organisational and socioeconomic predictor variables. Boosted regression trees combine elements of statistical boosting with regression trees. Regression trees use, in a first step, binary recursive splits to subdivide the domain of the various predictor variables into non-overlapping regions. In a second step, the reaction of match quality (the response variable) in every region is set to a constant region-specific value to minimise a loss function. In a third step, the predicted match quality is computed by aggregating over the regions (for a comprehensive analysis of regression trees, see Breiman et al. 1983). Regression trees capture potential interaction effects between predictor variables, they provide a unified platform to study different variable types (nominal, ordinal, cardinal), they use surrogate splits to model missing data, and they are robust to outliers and to the inclusion of irrelevant predictors (Breiman et al. 1983; Hastie et al. 2009; James et al. 2013). At the same time, however, their hierarchical structure makes regression trees sensitive to small disturbances of the data. Boosted regression trees overcome this data sensitivity by combining a large number of trees (Friedman 2002). Special techniques have been developed for boosted regression trees that render it possible to study in detail how match quality depends on a large group of organisational and socioeconomic predictor variables, how match quality responds to interactions of the predictor variables, and how important the various predictor variables are for predicting the cross-section of match quality across a large number of volunteers.

We proceed as follows. In Section 2, we describe how boosted regression trees are computed. In Section 3, we describe the data that we use to construct our index of match quality, our data on the Internet use of volunteers, and our data on several control variables. In Section 4, we summarise our empirical findings. In Section 5, we conclude.

2. Boosted Regression Trees

While boosted regression trees have been studied extensively in the machine learning literature since the pioneering work of Friedman (2001 and 2002), machine learning techniques have come to the attention of economists only very recently (for a survey, see Varian 2014). Empirical research in economics typically belongs to what Breimann (2001b) calls the "data modelling culture" in statistics, that is, the standard way of approaching an empirical research question is to assume a function (for example, a linear regression model) that links a response variable to predictor variables, to estimate the parameters of the function, and then to use the estimated function to predict the response variable. Boosted regression trees, in turn, belong to what Breimann (2001b) calls the "algorithmic modelling culture." Rather than starting with an assumption on the function that links a response variable to predictor variables, a data-driven algorithm (that is, a regression tree) is used to trace out during the modelling process the function that links the response variables to the predictors. A detailed exposition of boosted regression trees can be found in the textbook by Hastie et al. (2009, chapter 9). For an introduction, see the textbook by James et al. (2013, chapter 8).

A regression tree partitions the domain of the n predictor variables, $x = (x_1, x_2, \dots, x_n)$, into l disjoint regions, R_l. At the top level of a regression tree, these regions can be constructed given a predictor, $s = 1, 2, \dots, n$, and a partitioning point (split point), p, as the half-planes $R_1(s,p) = \{x_s | x_s \leq p\}$ and $R_2(s,p) = \{x_s | x_s > p\}$, where our notation follows the one used by Hastie et al. (2009, chapter 9). The optimal partitioning predictor and the optimal split point can then be identified using the following simple *Tree Algorithm*:

1. Loop over partitioning predictors: s in 1 to n:
   (a) Loop over all split points p:
      i. Construct half-planes: $R_1(s,p) = \{x_s | x_s \leq p\}$ and $R_2(s,p) = \{x_s | x_s > p\}$.
      ii. Compute for a squared error loss function: $\bar{y}_k = mean\{y_i | x_s \in R_k(s,p)\}$, where $k = 1, 2$ and y_i denotes observation i of the response variable.
   (b) Compute and store the loss for every split point: $L_{s,p} = \sum_k \sum_{x_s \in R_k(s,p)} (y_i - \bar{y}_k)^2$.

2. Select the optimal partitioning predictor and the optimal split point: $\min_{s,p} L_{s,p}$.

Once the first partitioning predictor and the first split point have been identified, the emerging regression tree has two terminal nodes. For this new tree, the *Tree Algorithm* is repeated separately for the optimal two half-planes identified at the top level. As this partitioning process continues, a complex hierarchical regression tree emerges that sends the predictors down its branches into the optimal regions, R_l. Once the partitioning process stops, the response variable (that is, match quality), can be predicted by checking the L terminal nodes of the tree:

$$T\left(\mathbf{x}_i, \{R_l\}_1^L\right) = \sum_{l=1}^{L} \bar{y}_l \mathbf{1}(\boldsymbol{x}_i \in R_l), \tag{1}$$

where T denotes a tree and **1** denotes the indicator function. Equation (1) shows that, even though in every particular region the prediction of match quality is simply given by its constant region-specific mean, the L region-specific means allow complex responses of match quality to the predictor variables to be modelled. Equation (1) also shows, however, that the hierarchical structure of a tree implies that if data are sent, say, to the left at an upper node of a tree then this allocation decision propagates through the entire remaining lower tree. A regression tree, thus, is sensitive to small changes in the data, making it a high-variance predictor. Among the techniques that have been advocated in the machine learning literature to remedy the data sensitivity of regression trees are bootstrap aggregation (Breiman 1996), random forests (Breimann 2001a), and boosted regression trees. These techniques have in common that the response variable is modelled by combining a large number of regression trees.

We grew regression trees in our empirical research using a boosting algorithm known as gradient least-squares boosting (Friedman 2001 and 2002). Retaining the assumption of a squared error loss function, $L = (y–F(\boldsymbol{x}))^2$, gradient least-squares boosting aims at minimising this loss function by approximating the unknown function, $F(\boldsymbol{x})$, which maps the predictors into predictions of match quality. Adopting a notation that follows the one used by Friedman (2001), the following *Boosting Algorithm* describes how to find such an approximation:

1. Initialise the algorithm: $T_0 = \bar{y}$.
2. Define some upper bound, M, for the number of iterations.
3. For m in 1 to M:
   (a) Compute the pseudo residuals: $\hat{y}_i = y_i – F_{m-1}(\boldsymbol{x}_i)$, where $i = 1 \ldots, N$.
   (b) Fit a regression tree, T_m, to the pseudo residuals to minimise L.
   (c) Update the function estimate: $F_m(\boldsymbol{x}) = F_{m-1}(\boldsymbol{x}) + T_m\left(\boldsymbol{x}, \{R_{l,m}\}_1^{L,m}\right)$.

(d) Equipped with the new function estimate, go back to Step (a).

4. Compute the final function estimate: $F(\boldsymbol{x}) = \sum_{m=0}^{M} T_m\left(\boldsymbol{x}, \{R_{l,m}\}_1^{L,m}\right)$.

The number of iterations, M, can be determined by first splitting the dataset into a training set and a prediction set, and then by searching for the minimum of the prediction error for alternative values of M. Alternatively, the optimal M can be fixed using cross validation. In order to scale down the influence of individual regression trees on the function approximation and, thus, to make the algorithm more robust, Friedman (2001) suggests to modify Step 3(c) in the *Boosting Algorithm* by introducing a learning rate, $0 < \zeta \leq 1$. Step 3(c) becomes

(2)
$$F_m(\boldsymbol{x}) = F_{m-1}(\boldsymbol{x}) + \zeta T_m\left(\boldsymbol{x}, \{R_{l,m}\}_1^{L,m}\right).$$

A smaller learning rate implies that more regression trees are needed to approximate $F(\boldsymbol{x})$. Hence, if the learning rate is fixed at a small value, the maximum number of iterations, M, should be large. In the boosting literature, it is common practice to choose $\zeta < 0.1$, and to set $M = 1,000$ or larger.

As a further modification, Friedman (2002) introduces a bootstrapping element into boosting. The resulting stochastic gradient boosting algorithm requires to sample without replacement, in every iteration, a subset from the data. Only the sampled data are then used to estimate the next tree, T_m, in the *Boosting Algorithm* in Step 3(b). Boostrapping should help to lower the correlation of predictions of individual trees and to make, as in a standard economic portfolio model that builds a low variance portfolio from uncorrelated assets, boosted regression trees a low-variance estimator.

3. The Data

After pre-sample testing based on interviews with 32 volunteers, an online questionnaire study was conducted from April to May 2013. The link to the online questionnaire study was distributed among GRC volunteers by means of a top-down snowball approach. Filling in the online questionnaire took approximately $20-30$ minutes of time. In total, $N = 4,611$ volunteers participated in the survey, resulting in a rich dataset.

It should be noted already at this stage of the analysis that by no means can we claim that this rich dataset represents the entire population of GRC volunteers. The GRC is a large organisation that consists of a national entity, 19 regional branches, and more than 420 district branches, and a national Federation of Nurses' Associations with 34 nurses' associations.[1] More than 400,000 volunteers work for the GRC.

[1] For more information on the GRC, see http://www.drk.de/ueber-uns/auftrag/english.html.

3.1 Measuring Match Quality

Table 1 shows the motive and utility dimensions for which we collected data to construct our match-quality index. For every volunteer in our dataset, we use data for 14 motive and utility dimensions. The motive and utility dimensions cover altruistic dimensions (public-goods model), egoistic motives like having fun and spending leisure time in a worthwhile manner (private-consumption model), and extrinsic dimensions related to job-market performance and social networking (human-capital model). Volunteers could rank separately every motive and every utility dimension on a 5-point scale from "do not agree" to "totally agree." Motives and utility reflect a volunteer's experiences at the time the survey study was undertaken (that is, volunteers were not asked to evaluate in retrospect their motives for becoming a volunteer).

Table 1

Motive and Utility Dimensions Used to Construct the Matching Index

Match dimension
to work together with other individuals
to contribute to small-scale developments in society
to bring about changes in politics
to improve one's standing within the GRC
to improve job-market prospects
to acquire job-market skills
to spend leisure time in a worthwhile manner
to improve one's standing in other parts of society
to come to other's attention within the GRC
to come to other's attention in society
to have fun
to help others
to strengthen the GRC
to defend one's interests

We compute a match-quality index (MQI) by aggregating the motive and utility dimensions using the following formula:

$$\text{MQI}_i = \sum_{j=1}^{14} \text{motive}_{i,j} \times \text{utility}_{i,j}, \qquad (3)$$

where i denotes a volunteer index. Thus, for every volunteer, we combine a motive dimension with a corresponding utility dimension in a multiplicative way and then form a volunteer-specific MQI by summing up over the 14 motive-utility dimensions for which we collected data. When both a motive dimension and the corresponding utility dimension receive a high rank then match quality is high along this dimension. Furthermore, the high rank implies that the respective dimension is important for a volunteer, and so this dimension automatically receives a large weight in the construction of the match-quality index. When both a motive dimension and a utility dimension receive a low rank then match quality is also high in this dimension, but the dimension automatically receives a small weight in the construction of the match-quality index because the dimension is comparatively less important for a volunteer. Finally, when a motive dimension receives a low rank but the corresponding utility dimension receives a high rank, this dimension receives a less than full weight because the differences in the ranks signal a mismatch along this dimension. Figure 1 plots the resulting match-quality index. After removing observations for which an MQI could not be computed due to missing data, data were available for $N = 3,527$ volunteers. The MQI has a mean of 152.09 index points, a median of 149 index points, and a standard deviation of 44.75 index points.

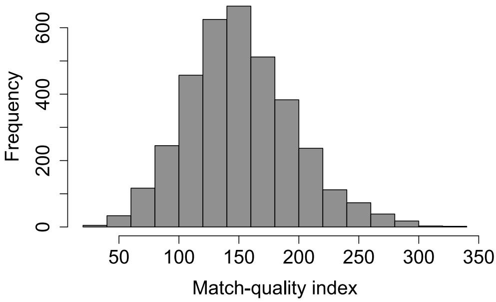


Figure1: Match-Quality Index ($N = 3,527$)

The MQI given in Equation (3) is constructed as suggested by Stukas et al. (2009), who compute a match-quality index using motives and affordances. The MQI has several interesting properties. First, the MQI can be easily calculated at the microeconomic level for every volunteer. Second, the MQI condenses different dimensions of motives and utility experiences into a single index. It is not necessary to define separate indexes for different motive and utility dimensions. In principle, constructing such indexes would be possible (see, for example, Emrich and Pierdzioch 2015), but in some cases it may be difficult to differentiate between, for example, altruistic motives and consumption motives. Third, Stukas et al. (2015) argue that their match quality index that combines motives and utility experiences (or, in their terminology, affordances) in a multiplicative way builds on well-established concepts and models widely used in psychology like the expectancy-value model, which predicts behavior by combining the subjective success probability with the subjective importance of a goal in a multiplicative way. Similarly, expected-utility models used in economics multiply the probability of success with the utility an agent can attain in a given state of the world. Finally, Stukas et al. (2015) show that their Total Match Index, which comprises the six elements of the so called Volunteer Functions Inventory (Clary et al. 1998), predicts volunteer outcomes (satisfaction, intention to volunteer in the future) better than than any univariate match index. Volunteer outcomes, in turn, are likely to be positively correlated with social capital.[2]

It is, thus, interesting to study whether our MQI correlates with measures of volunteer satisfaction. Table 2 shows summary statistics of various measures of volunteer satisfaction. We collected data on the overall satisfaction of volunteers with their volunteering at the GRC, their satisfaction with camaraderie in the GRC, their satisfaction with the appreciation of volunteer work in the GRC, their satisfaction with the management culture of the GRC, and their satisfaction with any payments they had received for their volunteer work.[3] Volunteers could rank their satisfaction on a 5-point scale, where higher values correspond to higher satisfaction levels.

Table 2

Measuree of Volunteer Satisfaction

Satisfaction measure	*N*	Mean	Median	SD
Satisfaction with volunteering	3,518	4.11	4.00	0.76
Satisfaction with comradeship	3,520	4.11	4.00	0.88
Satisfaction with appreciation	3,515	3.74	4.00	1.09
Satisfaction with management	3,497	3.31	3.00	0.95
Satisfaction with payments	1,242	3.45	4.00	1.14

Note: SD = standard deviation.

The results given in Table 3 show that the association of the match-quality index with the various dimensions of volunteer satisfaction is positive. Higher values of the match-quality index are on average accompanied by higher levels of volunteer satisfaction. The positive association between the match-quality index and volunteer satisfaction is in line with economic intuition insofar as

2 See Stukas et al. (2005), who find that measures of social capital are higher for "matched" volunteers.

3 84.4% of the respondents stated that they receive a compensation for expenses of less than 100 euros per month, 11.1% stated that they receive a compensation for expenses between 100 and 300 euros per month, 4.4% stated that they receive more than 300 euros per month.

one would expect that a low match quality results in low volunteer satisfaction, though volunteer dissatisfaction in principle can also result if match quality is high.

Table 3

Association of the Match-Quality Index With Measures of Volunteer Satisfaction

Rank of satisfaction measure	1	2	3	4	5
Satisfaction with volunteering	114.55	123.77	132.18	149.62	167.57
Satisfaction with comradeship	108.48	125.89	141.62	149.33	163.09
Satisfaction with appreciation	119.90	139.96	148.78	153.38	161.58
Satisfaction with management	121.18	137.03	149.77	159.71	170.25
Satisfaction with payments	156.61	149.04	159.49	159.30	164.97

Note: This table shows the mean value of the MQI for the different ranks of the satisfaction measures.

Finally, we mention that, despite similarities in terms of terminology, the way we measure match quality differs in many ways from how match quality is commonly measured in labour economics, where match quality is a key concept. In labour economics, match quality typically refers to time-invariant but initially unknown match-specific worker productivity. Match quality is an experience good because new information on worker productivity is revealed during a match (Jovanovic 1979). While widely studied classical search-and-matching models underscore the key importance of match quality, direct evidence on match quality is hardly available due to a lack of micro-level data on the productivity of a worker-firm match. In empirical research, data on job tenure and wages have been used to proxy match quality (see, for example, Bowlus 1995), but such an approximation is not useful in the case of volunteers because volunteers "work for nothing" (or, in many cases, they receive a rather low or more or less symbolic compensation for expenses). An alternative approach is to measure match quality using output data. For example, Jackson (2013) uses data on student test scores to measure how the quality of a match between a teacher and a school shapes student achievement. Such an output-based approach is hardly applicable because the German Red Cross mainly produces output for third parties that are not members of the GRC, and we do not have available data on, for example, the health condition of an accident victim rescued by a volunteer team of GRC emergency medical technicians. Yet another approach used in labour economics is to measure match quality using data on job satisfaction or job-related well-being (e.g. Freeman 1978; Akerlof et al. 1998; Green 2010). However, job satisfaction, while having the advantage that it also captures nonpecuniary rewards, is only an indirect measure of match quality, even though job satisfaction and match quality should exhibit a stronger positive correlation than match quality and wages (Gielen 2013). Moreover, job satisfaction often can be measured along several different dimensions (see also Table 2 for our data), giving rise to the question of how to condense this information into a single and comprehensive index of match quality (Barmby et al., 2012).[4]

3.2 Data on Internet Use

Table 4 summarises information on the intensity of Internet use. We ask volunteers to rate on a 5-point scale how intensely they use the Internet for GRC-related volunteering purposes and how intensely they use the Internet site of the GRC. In addition, we ask volunteers to rate on a 5-point scale how intensely they use the Internet in their leisure time. The three categories of Internet use are likely to overlap, but the model that we use in our empirical analysis traces out the partial effects of the three categories of Internet use on the MQI. The summary statistics depicted in Panel A show that volunteers use the Internet for GRC-related volunteering purposes more intensely than the Internet site of the GRC. Panel B shows that the MQI exhibits a positive association with all three categories of Internet use.

By means of binary variables we collect additional data to study which task volunteers perform using the Internet. Specifically, we ask volunteers to state whether they use the Internet for the following purposes: information collection, e-mail communication, use of social media, media consumption, and other purposes. Panel C of Table 4 reports summary statistics. Approximately 96 % of volunteers state that they use the Internet to collect information and to communicate via e-mail. Hence, these two predictor variables should hardly have any discriminatory power. The same can be expected for the variable "Internet used for other purposes." Because boosted regression trees are insensitive to the inclusion of irrelevant predictors, we keep these three variables in the model. Things are different for social-media use and media consumption. Around 60 % of the volunteers use the Internet to participate in social media, and only 32 % state that they use the Internet for media consumption. Out of the five categories, the use of the Internet for participation in social media shows the strongest association with the MQI.

Boosted regression trees as a technique representing a data-driven "algorithmic modelling culture" (Breimann 2001b) are particulary suited for our exploratory empirical analysis insofar as we do not begin our research with pre-

[4] For a proposal as how to combine data on job satisfaction to a measure of match quality, see Ferreira and Taylor (2011); for an index of volunteer satisfaction, see Galindo-Kuhn and Guzley (2001).

Table 4

Measures of Internet Use

Panel A: Primary Classification of Internet Use (5-Point Scale)

Dimension of internet use	*N*	Mean	Median	SD
Volunteer uses Internet for GRC volunteering	3,501	4.16	4.00	0.92
Volunteer uses Internet site of the GRC	3,505	3.37	4.00	1.15
Volunteer uses Internet during his / her leisure time	3,502	2.49	3.00	1.14

Panel B: Association of the Match-Quality Index (Mean Values) With Internet Use

Intensity of Internet use	1	2	3	4	5
Volunteer uses Internet for GRC volunteering	125.96	138.70	150.65	158.53	165.83
Volunteer uses Internet site of the GRC	137.57	149.24	156.63	164.07	179.79
Volunteer uses Internet during his / her leisure time	127.88	148.35	142.14	149.39	158.92

Panel C: Other Classification of Internet Use (Binary)

Dimension of internet use	%
Volunteer uses Internet to collect information	96
Volunteer uses Internet for e-mail communication	97
Volunteer uses Internet to participate in social media	60
Volunteer uses Internet for media consumption	32
Volunteer uses Internet for other purposes	08

Note: SD = standard deviation.

cise (theory-based) hypotheses as to how the various forms of Internet use may affect match quality. For example, use of social media may improve the social integration of users and may, thereby, help to improve match quality. Use of social media, however, may also be unrelated to match quality or it may even deteriorate match quality if using social media means that volunteers invest less time in their search for a volunteer position that matches their motives. Similarly, match quality may improve if volunteers use the Internet in their leisure time. Alternatively, match quality may deteriorate if a volunteer uses the

Internet intensively for leisure-time activities because the resources spent on optimising GRC-related volunteering activities become scarcer. Similarly, we do not start with specific hypotheses as to how and – if so – to which extent Internet use interacts with the other control variables that we study in our empirical research.

3.3 Control Variables

Panel A of Table 5 shows summary statistics of GRC-specific and volunteering-related control variables. We ask volunteers whether they have a contact person on site, whether the GRC supports them, whether they can participate in decision-making processes, and whether they attended training courses (on volunteer management and volunteer satisfaction, tenure, and commitment, see Cnaan and Cascio 1999). Training programmes could help to recover information about a volunteer's match-specific productivity. At the same time, training periods could help a volunteer to "learn-on-the-job." In the labour economics literature, learning-on-the-job effects typically are differentiated from match-quality effects (see Nagypal 2007). In case of our match-quality index, it is difficult to disentangle the effect of learning-on-the-job and match quality on the congruence of motives and utility, and given a lack of data that could be used for this purpose we make no attempt in this direction.

Match quality may improve if a volunteer can choose when to do volunteer work. Also, match quality may improve if a volunteer has done volunteer work for the GRC before he or she entered into the current volunteer position (for a search model of branch and firm search, see Neal 1999). Match quality could also be linked to the duration of the current volunteer position. On the one hand, match quality could improve in duration because a long-serving volunteer could be familiar with how the GRC is organised and with the requirements of a volunteer position. A long-serving volunteer, thus, may have had the chance to find out whether the requirements of a volunteer position match his or her motives for doing volunteer work. As a result, classical search models of the labour market predict that match quality increases with tenure (see Jovanovic 1979). On the other hand, job-specific human capital lowers the incentive to search for another volunteer position within the GRC or in another volunteer organization. In the terminology of labour economics, a long-serving volunteer who has accumulated much job-specific human capital will reject "outside offers" even if match quality is low (see Barmby et al. 2012 and the references cited therein).

We also ask volunteers whether their volunteer position can be characterised as a management position (director, head of a unit, treasurer, assessor), an executive position (physician, media-relations representative, in-house counsel, technician, trainer, minutes secretary), or another position. Because a volunteer can be a member of the board of management of his or her organisational GRC unit and, at the same, work as a technician in this unit, the proportions of management, executive, and other positions do not sum up to unity. Our definition of a "management position" and an "executive position" is admittedly arbitrary but, as we shall show in Section 4.2, the relative importance of these two control variables – and therefore their impact on our empirical results – is negligible.

Next, we ask volunteers whether they were elected to their volunteer position. An election can be interpreted as an active search process and, at the same time, may result in a specific loyality bond to the organisational unit in which a volunteer works or even to the GRC. Finally, the size of a GRC unit for which a volunteer works and the proportion of volunteers in this unit may matter because a free-rider problem may beleaguer large units, so that a volunteer does a job simply because no one else wants to do the job, which could result on average in a low match quality. The size of a GRC unit, however, also may have a positive effect on match quality because more volunteer positions are available in larger organisational units.

Panel B of Table 5 shows summary statistics of further socioeconomic control variables. We control for a potential gender effect and the proportion of friends in the GRC. A large proportion of friends in the GRC may leverage self-reports of match quality because having more friends in an organisation may alleviate the gradual revelation of information on match-specific productivity. At the same time, a large proportion of friends in the GRC may be negatively correlated with match quality if a volunteer does not change a volunteer position with a low match quality only because a volunteer wants to continue working with his or her friends in a GRC unit. Furthermore, match quality may be high if volunteering plays an important role in a volunteer's life. At the same time, if volunteering plays an important role in a volunteer's life, self-deception may require to bring utility experiences in line with motives for volunteering. Finally, to capture general humanitarian values of a volunteer we ask about interest in politics and religiosity. A stronger religiosity, for example, may improve match quality because a volunteer may increase efforts to find the best position within the GRC. At the same time, however, stronger religiosity may imply that a volunteer is more interested in participating in religious organisations such that match quality could decrease.

Table 5 shows that data on the predictors are not available for all volunteers who participated in the questionnaire study, underlining the fact that the *Boosting Algorithm* automatically deals with missing data is of great value for our empirical analysis. Table 5 further shows that we study different variable types. For our analysis, another advantage of the algorithm is that it provides a modelling platform for the analysis of different variable types. Furthermore, a detailed modelling of the differential effects of the various binary and ordinal control variables would require definition of more than thirty dummy variables in the context of a classical regression model. Another advantage of the *Boosting*

Table 5

Control Variables

Panel A: GRC and Volunteer-Related Control Variables

Control variable	N	Mean	Median	SD	Description
Contact person	3,505	–	–	–	contact person available (no = 0, yes = 1, 58\ % answered yes)
Support	3,501	3.46	3.00	0.92	support by the GRC (5-point scale from never to always)
Participation	3,511	3.51	4.00	1.09	volunteer can take part in decision making (5-point scale)
Training	2,813	1.81	2.00	0.49	volunteer attended training courses (3-point scale, no = 0, one course = 1, more courses = 2)
Size	2,032	4.91	4.38	1.77	size of GRC organisational unit for which a volunteer works (in logs)
Flexibility	3,515	3.62	4.00	1.02	volunteer can choose when to do volunteer work (5-point scale)
Volunteered before	3,514	–	–	–	volunteer volunteered for the GRC before (no=0, yes=1, 37% answered yes)
Duration	3,472	11.22	7.00	10.79	how long has the volunteer worked for the GRC (in years, less than a year = 0)
Elected	3,497	–	–	–	volunteer was elected to position (no=0, yes=1, 45% answered yes)
Management position	3,527	–	–	–	volunteer has a management position (43% answered yes)
Executive position	3,527	–	–	–	volunteer has an executive position (41% answered yes)
Other position	3,527	–	–	–	volunteer has another position (not management or executive, 50% answered yes)
Proportion	3,208	56.13	60.00	34.49	proportion of volunteers in the organisational GRC unit for which a volunteer works (in percent)

Panel B: Other Socioeconomic Control Variables

Control variable	N	Mean	Median	SD	Description
Friends	3,518	2.85	3.00	1.08	proportion of friends working for the GRC (5-point scale from low to high)
Sex	3,493	–	–	–	male = 0, female = 1 (35% females)
Importance of volunteering	3,521	4.15	4.00	0.78	overall importance of volunteering (5-point rank from unimportant to very important)
Political interest	3,503	3.83	4.00	0.87	interest in politics (5-point scale from uninterested to very interested)
Religiosity	2,472	2.72	3.00	1.04	religiosity (5-point scale from unimportant to very important)

Note: SD = standard deviation.

Algorithm is that we do not have to specify before estimation of the model interaction effects between the control variables, nor do we have to decide *a priori* on the functional form of potential nonlinear effects. Finally, the *Boosting Algorithm* has the advantage that any variables on our list of control variables that have weak explanatory power for match quality are hardly used as splitting variables, implying that such variables have a negligible relative importance for the overall fit of the model.

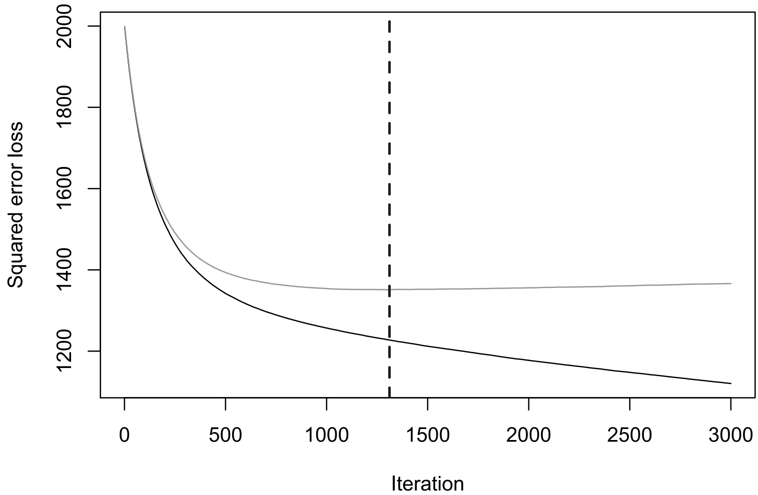
4. Empirical Results

4.1 Model Specification

Estimation of boosted regression trees requires calibration of a few parameters. Two key parameters are the learning rate and the total number of trees. A smaller learning rate implies that, as Equation (2) shows, more trees are needed to compute a good approximation of the predictor function, $F(\boldsymbol{x})$. Adding more trees, however, involves a trade-off. On the one hand, the boosted regression trees start tracking more closely even a complicated predictor function. Any estimation bias that results when too few trees are estimated gets smaller. On the other hand, the variance of the predictor increases as the model starts picking up any idiosyncracies of the estimation data. The predictive power of the model when applied to new or somewhat disturbed data then is likely to be poor, that is, an overfitting problem arises. The resulting trade-off between variance and bias implies that it is possible to detect an optimal number, M^*, of trees for a given learning rate.

Figure 2 illustrates the resulting trade-off for a learning rate of $\zeta = 0.005$. The black line shows that adding more trees results in a successively smaller squared error loss computed for the estimation data. The squared error loss for the test data (grey line; in this research, we compute the test data using 5-fold cross-validation) falls due to the bias reduction as additional trees are added to a small-scale model. Adding more trees, however, also inflates variance, and so the squared error loss for the test data starts increasing at some point as more and more trees are added to an already large model. The model features an optimal number, M^*, of trees for the given learning rate when the effects of bias reduction and variance inflation on the squared error loss balance. In Figure 2, the optimal number of trees is $M^* = 1,311$ (dashed vertical line).

Another parameter that has to be calibrated is the tree size. The tree size governs the maximum number of terminal nodes per tree. Setting tree size to one gives stumps and so the resulting additive model would neglect any potential interaction effects between the predictors. In order to capture potential interactions of the predictor variables in general and of the control variables with Internet use in particular, we assume a tree size of 5. Assuming a tree size of 10



Note: Grey line = squared error loss for the test data obtained by means of 5-fold cross-validation. Black line = squared error loss computed for the estimation data. Dashed vertical line = optimal number of trees. Learning rate = 0.005. Test and training data were obtained by 5-fold cross-validation, minimum number of observations per terminal node = 10. Tree size = 5. Boosting proportion = 0.5.

Figure 2: Determination of the Optimal Number of Trees

leads to similar results. Furthermore, we fix the minimum number of observations per terminal node to assume the value 10. Finally, as suggested by Friedman (2002), we add an element of stochasticity to our boosted regression trees by selecting at random 50 % of the data in every iteration of the *Boosting Algorithm*. Increasing the proportion of randomly selected data to 75 % yields similar results (not reported).

We compute all empirical results documented in this research using the R programming environment for statistical computing (R Core Team 2015) to carry out our empirical analysis. For estimation of boosted regression trees, we use the add-on package "gbm" (Ridgeway 2015). Our data feature missing predictor data because not all volunteers answered all questions. The "gbm" package handles missing data on predictor variables using surrogate splits (for the case of a classifier model, see Breimann et al. 1983, 142).

4.2 Relative Influence of Predictor Variables

Equipped with the calibrated parameters, we estimate the boosted regression trees and study the relative influence of the various predictor variables. Relative influence is defined as the improvement in the squared error resulting from using a predictor to form splits (Breimann et al. 1983), averaged across base learners (Friedman 2001). Due to the stochasticity of the algorithm, Table 6 summarises the mean and the standard deviation of the relative influence of the predictor variables as obtained by running simulations of the model. Numbers are in percent. For example, using the predictor "Participation" on average improves squared error by 6.38 %.

Table 6

Relative Influence of Predictor Variables

Predictor Variable	Mean	SD
Importance of volunteering	41.19	0.94
Support	11.01	0.21
Participates in social media	8.60	0.22
Participation	6.38	0.10
Friends	4.12	0.09
Proportion of volunteers	3.62	0.21
Flexibility	3.44	0.16
Duration	3.23	0.29
Internet of GRC	3.19	0.09
Internet for GRC	2.95	0.09
Size	2.89	0.24
Political interest	2.66	0.18
Internet used in spare time	1.73	0.13
Religiosity	1.18	0.10
Elected	1.10	0.13
Training attended	0.87	0.07
Internet used for media consumption	0.50	0.04
Contact person	0.39	0.06
Sex	0.32	0.05
Executive position	0.26	0.05
Volunteered before	0.15	0.04
Management position	0.07	0.02
Other position	0.06	0.02
Internet used for other purposes	0.03	0.01
Internet used for e-mail communication	0.02	0.01
Internet used to collect information	0.02	0.01

Note: Relative influence (measured in percent) is defined as the improvement in squared error resulting from using a predictor to form splits, averaged across base learners. Mean = mean relative influence. SD = standard deviation of relative influence. Results are based on 250 model simulation runs (50% sampling rate, learning rate 0.005, 5-fold cross-validation, mean number of optimal base learners = 1,279.88, SD of the optimal number of base learners = 96.66).

In terms of relative influence, the subjective importance of volunteering as perceived by a volunteer is the most important predictor of match quality followed by the GRC-specific variables support and participation. Other predictor variables assuming positions in the upper part of the table are duration (that is, how long a volunteer has already worked in his or her current position), the flexiblity of the working conditions, and the proportion of volunteers in an organisational GRC unit. The relative influence of these predictor variables, however, is relatively small, and they are not among the top three predictor variables. As for the Internet variables, the top predictor variable is participation in social media, followed by the intensity of Internet use for GRC volunteering and the intensity of use of the Internet site of the GRC for volunteering – the latter two predictor variables have a substantially smaller relative influence than the former, however. The total relative influence of all Internet variables is 17 % (SD = 0.16).

As a robustness check, we changed the calibration of the parameters of the model. One key parameter is the learning rate. Increasing the learning rate from 0.005 to 0.1 substantially reduces the number of trees that form a BRT model, but the ordering of variables and the magnitude of the numbers are close to the results reported in Table 6 (results are not reported, but are available upon request). As a further check of the robustness of the model, we split the data into an in-sample subsample and an out-of-sample subsample, where the latter contained 20 % of the data. We then reestimate the boosted regression trees on the in-sample subsample and compute coefficients of determination for both subsamples. We find for the in-sample subsample that $R^2 = 0.44$ and for the out-of-sample subsample that $R^2 = 0.29$ (the coefficients of determination fluctuate to some extent randomly because of the stochasticity of the model). The empirical model, thus, has a satisfactory explanatory power and the out-of-sample R^2 underscores that it is not severely sensitive to perturbations of the data.

The BRT model detects key patterns in the data and, thus, is an ideal modelling platform for exploratory data analysis. The results of such an analysis can be used in various ways. One way is to use the results to develop a simple and tractable model of match quality. For example, the results summarised in Table 6 inform about the key predictor variables of match quality. The top three predictor variables are the subjective importance of volunteering, the GRC-specific variable support, and participation in social media. Taken together, the relative importance of these three predictor variables exceeds 60 %. This information can be used to set up a linear regression model. Such a linear regression model has its limitations and is much less flexible than the BRT model, but parsimony certainly is its key advantage when it only features the top three predictor variables identified by the BRT model. We estimate such a parsimonious linear regression model by least squares on the full dataset and find, as expected, that the top three predictor variables identified by the BRT model have significant explanatory power for match quality ($N = 3,485$,

F-test = 169.8 with a p-value of < 0.02), where the in-sample R^2 assumed a value of approximtely 0.30 (detailed estimation results are available upon request from the authors).

4.3 Partial Dependence

The partial dependence plots summarised in Figure 3 show how exactly the match quality index (vertical axis) correlates with several predictor variables (horizontal axis). The effects of the other predictors are taken into account using the weighted traversal technique described by Friedman (2001).

Match quality increases if a volunteer states that volunteering is an important part of his or her life and if a large proportion of a volunteer's friends also volunteer for the GRC. Match quality is higher on average if a volunteer has already worked for some time in his or her current position, but for a longer duration match quality decreases again.[5] Match quality increases if a volunteer receives support from the GRC and is involved in decision-making processes. Match quality tends to be lower if the proportion of volunteers in an organisational GRC unit is below approximately 20 %, but then increases rapidly. A small proportion of volunteers in an organisational unit may reflect that only relatively few positions are available for volunteers in such a unit, implying that a volunteer can work in relatively few different areas and / or on few different tasks. As a result, it may be harder to match volunteers to volunteer positions in such organisational GRC units. At the same time, match quality tends to decrease for intermediate-sized organisational GRC units. Match quality also increases in flexibility (that is, if a volunteer can decide when to do his or her work).

Match quality is positively correlated with the three Internet-related predictor variables: social-media usage, the intensity of usage of the Internet for GRC volunteering, and the intensity of usage of the Internet site of the GRC. As compared to social-media usage, however, the latter two predictors have a smaller relative influence. The positive correlation of match quality with the intensity of usage of two GRC-related forms of Internet use corroborates findings of Emrich and Pierdzioch (2016). They find that measures of volunteer commitment positively correlate with volunteers' use of the Internet for GRC volunteering and / or use of the Internet site of the GRC. Bringing together their results with the findings reported in Figure 3 suggests that – in the sample of volunteers studied in this research – one reason why intense Internet use correlates with volunteers' commitment is that it is associated with improved match quality.

The positive correlation of match quality with use of social media is consistent with recent results reported by Farrow and Yuan (2011). They analyse a sample of alumni of an American university and find that participation in social

[5] Compare also the results reported by Dekimpe and Degraewe (1997).

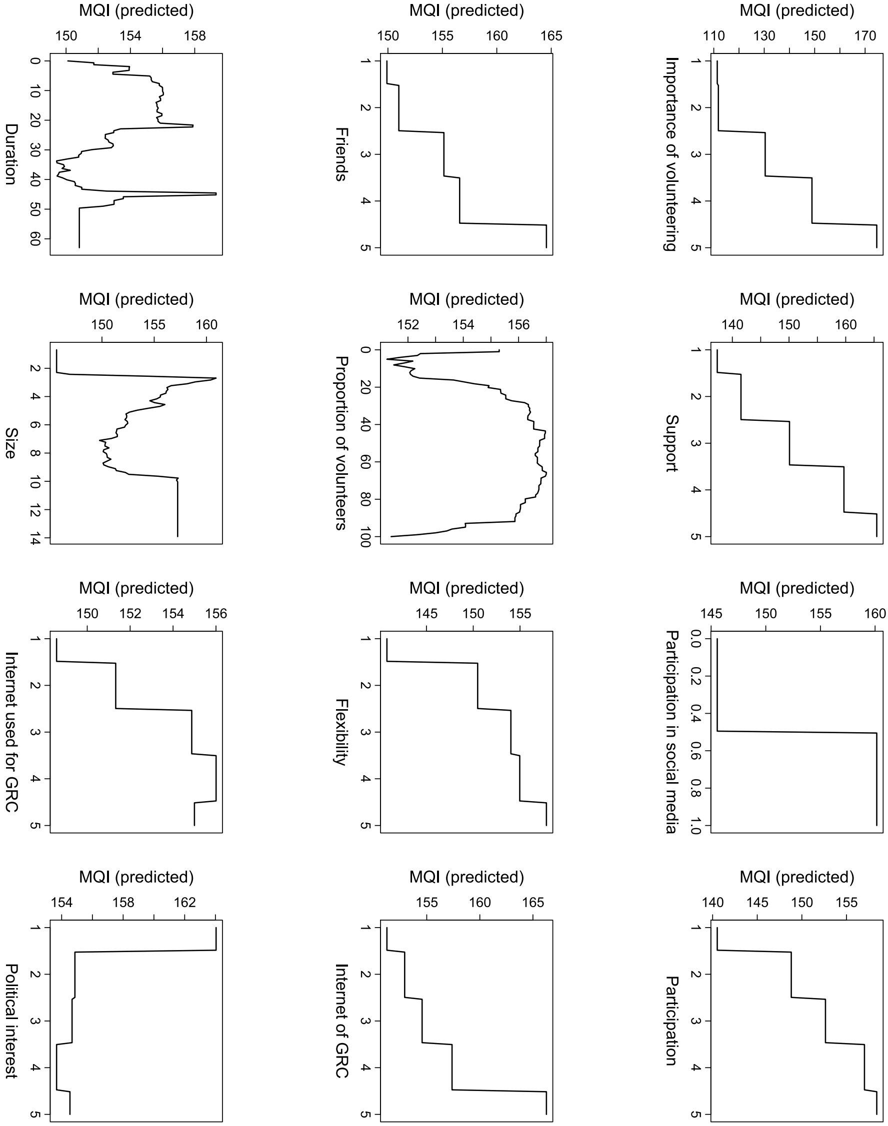


Figure: 3 Partial Dependence Plot

media groups (Facebook) strengthens network ties through frequency of communication with other alumni and emotional closeness to other alumni and to the university.[6] Emotional closeness in particular strengthens a positive attitude toward volunteering and charitable giving for the university, and both emotional closeness and frequency of communication eventually lead to actual volunteering and charitable giving.

Sabatini and Sarracino (2014) study Italian data and find that social-media use has a negative effect on social trust (probably because the radius of communication increases), but a positive effect on face-to-face communication. Results of extensive research on volunteering, in turn, demonstrate that face-to-face communication is an important aspect of volunteer recruitment, one reason being that individuals learn about opportunities to volunteer.[7] When a prospective volunteer has many diverse opportunities to volunteer, match quality is likely to increase. This effect may persist after recruitment.

The case for a positive correlation betwen social media use and social capital is strengthened by results reported by Steinfeld et al. (2008) and Ellison et al. (2008), who report that Facebook use helps users to form and maintain social capital, an effect that seems to be stronger for those users with low self-esteem. Similarly, Valenzuela et al. (2009) report a positive correlation between the intensity of Facebook use and life satisfaction, social trust, civic engagement, and political participation for a sample of college students. Results reported by de Zũniga et al. (2012) corroborate this positive correlation insofar as they find that usage of social networks for information purposes is positively associated with online and offline civic and political participation.

4.4 Interaction Effects

Internet use also affects how the other predictor variables are linked to match quality. Figure 4 illustrates such interaction effects for the GRC-related variables. The black line represents the partial dependence function of the GRC-related variables when a volunteer hardly uses the Internet for GRC-related purposes, while the grey line illustrates the resulting partial dependence function when a volunteer heavily uses the Internet for GRC-related purposes. Figure 4 illustrates how the switch from low to high intensity of Internet use shifts the partial dependence functions upward. Such an upward of shift is intuitively plausible insofar as it can be interpreted to indicate that a higher intensity of Internet use for GRC-related purposes implies, for examle, that having a contact person within the GRC, receiving support from the GRC, and attending

6 For a definition of the strength of network ties, see Granovetter (1973). On the predictive power of social media use for tie strength, see Gilbert and Karahalios (2009).

7 For useful literature surveys, see Wilson (2000 and 2012).

training courses become more effective means of improving match quality. The key message to take home from Figure 4, thus, is that volunteers' Internet use has a direct effect on match quality, but it also has an indirect effect that propagates through the hierarchical structure of the boosted regression trees.

5. Concluding Remarks

Our results suggest that match quality, on the one hand, and the intensity of volunteering-related Internet use and especially social-media use, on the other hand, are positively correlated. Correlation does not establish causality because a high match quality could improve volunteers' motivation, and highly motived volunteers then may decide to use the Internet or to contact their friends via social media to search for additional information regarding training programmes or other GRC activities.[8] Hence, the interpretation of our results in terms of causal links running from Internet use to match quality should not be stretched too far. Still, the results of the partial dependence analysis suggest that Internet use does not necessarily crowd out civic engagement, and that it does not eventually destroy social capital. In this respect, it should also be noted that the empirical model recovered positive interaction effects between Internet use and exogenous predictor variables like participation opportunities, GRC support, availability of training programmes, and others.

A question that we could not analyse with our sample of data is how Internet use is intertwined with the flow into and out of the volunteer labour force. The sample that we analysed in this research consisted of active volunteers only, and so our results have nothing to say about the role Internet use plays for the recruitment of volunteers. By the same token, the results of the empirical analysis shed light on how match quality and Internet use are interrelated in the group of volunteers but do not inform about potential differences in the form and the intensity of Internet use across volunteers and non-volunteers. In this respect, it is also important to note that the positive correlation between Internet use and match quality detected by the boosted regression trees may in part be due to a survival bias because the sample of active volunteers analysed in this research, by construction, does not feature dissatisfied volunteers who terminated their volunteer work due to a low match quality.

Future research should also focus in detail on how exactly match quality is linked to Internet use. The intensity of Internet use in general and the use of social media in particular could strengthen the social ties between volunteers, resulting in a better flow of information on opportunities to volunteer. At the same time, the interaction of Internet use with participation opportunities de-

---

[8] On the endogeneity problem, see Bauernschuster et al. (2014) and Sabatini and Sarracino (2014).

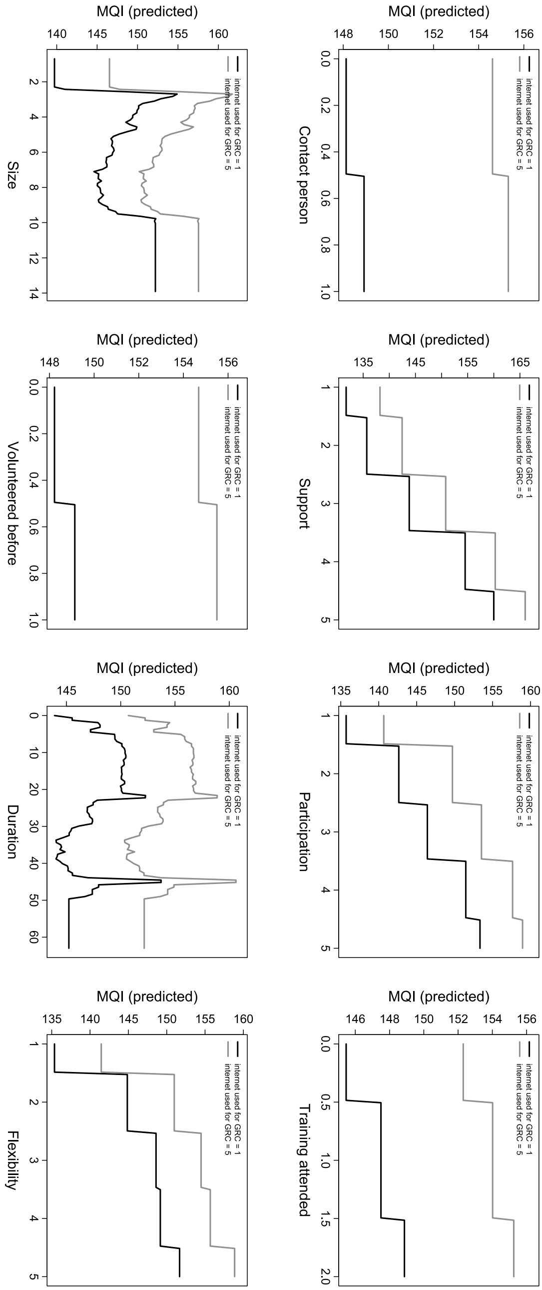


Figure 4: Interaction Effects

tected by the boosted regression trees suggests that the Internet may also be an effective communication medium that better integrates volunteers into internal communication processes within the GRC and that helps to perpetuate the exchange of information between GRC staff and volunteers. As Emrich and Pierdzioch (2016) emphasise, trust-dependent transaction costs may be important in this respect. Information and search costs are one element of trust-dependent transaction costs; conflict and enforcement costs, and surveillance and monitoring costs are other elements. The Internet can help to reduce such surveillance and monitoring costs by making internal decision processes and the internal allocation of resources more transparent, and it can help to enforce norms and rules of "good practice."[9] A by-product of transparency and the enforcement of norms could be that it becomes easier for volunteers to bring in line their utility experiences with their motives for doing volunteer work, such that not only match quality would increase but eventually also volunteers' commitment and satisfaction.

References

Akerlof, G., A. Rose, J. Yellen, L. Hall, and R. Hall. 1988. "Job Switching and Job Satisfaction in the U.S. Labor Market." *Brookings Papers on Economic Activity* 2: 495 – 594.

Andreoni, J. 1989. "Giving With Impure Altruism: Applications to Charity and Ricardian Equivalence." *Journal of Political Economy* 97: 1447 – 1458.

— 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *Economic Journal* 100: 464 – 477.

Barmby, T., A. Bryson, and B. Eberth. 2012. "Human Capital, Matching and Job Satisfaction." *Economics Letters* 17: 548 – 551.

Bauernschuster, S., O. Falck, and L. Woessmann. 2014. "Surfing Alone? The Internet and Social Capital: Evidence From an Unforeseeable Technological Mistake." *Journal of Public Economics* 117: 73 – 89.

Bergstrom, T., L. Blume, and H. Varian. 1986. "On the Private Provision of Public Goods." *Journal of Public Economics* 29: 25 – 49.

Boezeman, E. J. and N. Ellemers. 2008. "Pride and Respect in Volunteers' Organizational Commitment." *European Journal of Social Psychology* 38: 159 – 172.

Bowlus, A. J. 1995. "Matching Workers and Jobs: Cyclical Fluctuations in Match Quality." *Journal of Labor Economics* 13: 335 – 350.

Breiman, L. 1996. "Bagging Predictors." *Machine Learning* 24: 123 – 140.

[9] For analyses of the applicability of principle-agent theory in the context of the management of non-profit organisations, see Caers et al. (2006) and Coule (2015). For why ex-post expropriation risk explains why non-profit organisations exist in the first place, see Glaeser and Shleifer (2001).

Breiman, L. 2001a. “Random Forests.” *Machine Learning* 45: 5–32.

— 2001b. “Statistical Modeling: The Two Cultures.” *Statistical Science* 16: 199–215.

Breiman, L., J. H. Friedman, R. Oshen, and C. Stone. 1983. *Classification and Regression Trees*. New York: Chapman & Hall.

Caers, R., C. du Bois, M. Jegers, S. De Gieter, C. Schepers, and R. Pepermans. 2006. “Principal-Agent Relationships on the Stewardship-Agency Axis.” *Nonprofit Management and Leadership* 17: 25–47.

Clary, E. G., M. Snyed, R. D. Ridge, J. Copeland, A. A. Stukas, J. Haugen, and P. Miene. 1998. “Understanding and Assessing the Motivations of Volunteers: A Functional Approach.” *Joumal of Personality and Social Psychology* 74: 1516–1530.

Cnaan, R. and T. Cascio. 1999. “Performance and Commitment: Issues in Management of Volunteers in Human Service Organizations.” *Journal of Social Service Research* 24: 1–37.

Coule, T. C. 2015. “Nonprofit Governance and Accountability: Broadening the Theoretical Perspective.” *Nonprofit and Voluntary Sector Quarterly* 44: 75–97.

de Zúñiga, H. G., N. Jung, and S. Valenzuela. 2012. “Social Media Use for News and Individuals’ Social Capital, Civic Engagement and Political Participation.” *Journal of Computer-Mediated Communication* 17: 319–336.

Dekimpe, M. G. and Z. Degraewe. 1997. “The Attrition of Volunteers.” *European Journal of Operational Research* 98: 37–51.

Duncan, B. 1999. “Modeling Charitable Contributions of Time and Money.” *Journal of Public Economics* 72: 213–242.

Ellison, N. B., C. Steinfeld, and C. Lampe. 2008. “The Benefits of Facebook ‘Friends:’ Social Capital and College Students’ Use of Online Social Network Sites.” *Journal of Computer-Mediated Communication* 12: 1143–1168.

Emrich, E. and C. Pierdzioch. 2015. “Testing Economic Models of Volunteer Labor Supply: Some Empirical Evidence for the German Red Cross.” *Applied Economics* 47: 4247–4259.

— 2016. “The Internet and the Commitment of Volunteers: Empirical Evidence for the German Red Cross.” *Nonprofit and Volunatry Sector Quarterly* (forthcoming).

Emrich, E., C. Pierdzioch, and M. Oestmann. 2014. “The Intensity of Internet Use By Volunteers: Empirical Results for the Internet Portal of the German Football Association.” *European Sport Management Quarterly* 14: 238–258.

Farrow, H. and Y. C. Yuan. 2011. “Building Stronger Ties With Alumni Through Facebook to Increase Volunteerism and Charitable Giving.” *Journal of Computer-Mediated Communication* 16: 445–464.

Ferreira, P. and M. Taylor. 2011. “Measuring Match Quality Using Subjective Data.” *Economics Letters* 113: 304–306.

Franzen, A. 2003. “Social Capital and the Internet: Evidence from Swiss Panel Data.” *Kyklos* 56: 341–360.

Freeman, R. 1978. “Job Satisfaction as an Economic Variable.” *American Economic Review* 68: 135–141.

Friedman, J. H. 2001. “Greedy Function Approximation: A Gradient Boosting Machine.” *Annals of Statistics* 29: 1189–1232.

— 2002. “Stochastic Gradient Boosting.” *Computational Statistics and Data Analysis* 38: 367–378.

Galindo-Kuhn, R. and R. M. Guzley. 2001. “The Volunteer Satisfaction Index: Construct Definition, Measurement, Development, and Validation.” *Journal of Social Service Research* 28: 45–68.

Garner, J. T. and L. T. Garner. 2009. Volunteering an Opinion: Organizational Voice and Volunteer Retention in Nonprofit Organizations. *Nonprofit and Voluntary Sector Quarterly* 40: 813–828.

Gielen, A. 2013. “Repeated Job Quits: Stepping Stones or Learning About Quality?” *IZA Journal of European Labor Studies* 2: 7–22.

Gilbert, E. and K. Karahalios. 2009. “Predicting Tie Strength With Social Media.” *Proceedings of the Conferece on Human Factors in Computing Systems (CHI’09)*: 211–220.

Glaeser, E. L. and A. Shleifer. 2001. “Not-for-profit Entrepreneurs.” *Journal of Public Economics* 81: 99–115.

Granovetter, M. S. 1973. “The Strength of Weak Ties.” *American Journal of Sociology* 78: 1360–1380.

Green, F. 2010. “Well-Being, Job Satisfaction and Labour Mobility.” *Labour Econmics* 17: 897–903.

Hampton, K. and B. Wellman. 2003. “Neighboring in Netville: How the Internet Supports Community and Social Capital in a Wired Suburb.” *City and Community* 2: 277–311.

Harbaugh, W. T. 1998. “What Do Donations Buy? A Model of Philanthropy Based on Prestige and Warm Glow.” *Journal of Public Economics* 67: 269–284.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. New York: Springer.

Hustinx, L. 2010. “I Quit, Therefore I Am? Volunteer Turnover and the Politics of Self-Actualization.” *Nonprofit and Voluntary Sector Quarterly* 39: 236–255.

Jackson, C. K. 2013. “Match Quality, Worker Productivity, and Worker Mobility: Direct Evidence.” *Review of Economics and Statistics* 95: 1096–1116.

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. *An Introduction to Statistical Learning: With Applications in R*. New York: Springer.

Jovanovic, B. 1979. “Job Matching and the Theory of Turnover.” *Journal of Political Economy* 87: 972–990.

Menchink, P. L. and B. A. Weisbrod. 1987. “Volunteer Labour Supply.” *Journal of Public Economics* 32: 159–183.

Musick, M. A. and J. Wilson. 2008. *Volunteers: A Social Profile*. Bloomington: Indiana University Press.

Nagypal, E. 2007. "Learning by Doing vs. Learning about Match Quality: Can We Tell Them Apart?" *Review of Economic Studies* 74: 537–566.

Neal, D. 1999. "The Complexity of Job Mobility Among Young Men." *Journal of Labor Economics* 17: 237–261.

Putnam, R. D. 1995. "Bowling Alone. America's Declining Social Capital." *Journal of Democracy* 6: 75–78.

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Ridgeway, G., with contributions from others (2015). *gbm: Generalized Boosted Regression Models*. R package version 2.1.1. http://CRAN.R-project.org/package=gbm.

Roberts, R. D. 1984. "A Positive Model of Private Charity and Public Transfers." *Journal of Political Economy* 92: 136–148.

Sabatini, F. and F. Sarracino. 2014. "Will Facebook Save or Destroy Social Capital? An Empirical Investigation Into the Effect of Online Interactions on Trust and Networks." MPRA Working Paper Paper No. 53325. Available at http://mpra.ub.uni-muenchen.de/53325/. Accessed 21 June 2015.

Schiff, G. 1980. *Charitable Giving and Government Policy: An Economic Analysis*. Westport, Connecticut: Greenwood Press.

Steinfeld, C., N. B. Ellison, and C. Lampe. 2008. "Social Capital, Self-Esteem, and Use of Online Social Networks: A Longitudinal Analysis." *Journal of Applied Development Psychology* 29: 434–445.

Stukas, A. A., M. Daly, and M. J. Cowling. 2005. "Volunteerism and Social Capital: A Functional Approach." *Australian Journal on Volunteering* 10: 35–44.

Stukas, A. A., K. A. Worth, E. G. Clary, and M. Snyder. 2009. "The Matching of Motivations to Affordances in the Volunteer Environment – An Index for Assessing the Impact of Multiple Matches on Volunteer Outcomes." *Nonprofit and Voluntary Sector Quarterly* 38: 5–28.

Valenzuela, S., N. Park, and K. F. Kee. 2009. "Is There Social Capital in a Social Network Site? Facebook Use and College Students' Life Satisfaction, Trust and Participation." *Journal of Computer-Mediated Communication* 14: 875–901.

Wilson, J. 2000. "Volunteering." *Annual Review of Sociololgy* 26: 215–240.

— 2012. "Volunteerism Research: A Review Essay." *Nonprofit and Voluntary Sector Quarterly* 41: 176–212.