http://dx.doi.org/10.15830/amj.2017.19.1.35

Peer Effects in Service Usage*

Sangyoung Song**

Researchers in marketing, sociology, and economics have been interested in the role of social interactions in consumer choice and consumption behaviors. Social interactions, labeled variously as peer effects, social contagion, and neighborhood effects, have important implications for firms' allocation of marketing efforts. In this research, we test and provide empirical evidence for peer effects in consumers' service usage in the context of gym-going behaviors. Using a detailed individual-level membership and attendance data at one of the largest health club chains in the U.S., we document that a focal member's gym-going behavior is influenced by the behaviors and characteristics of the peers at the same branch.

Key words: Peer effects; Social Contagion; Quantile Regressions, Instrumental Variables

Ⅰ. Introduction

Social contagion has been studied under a variety of names; bandwagon effects, peer effects, neighborhood effects, and interdependent preferences (Glaeser, Sacerdote, and Scheinkman 2003; Lyle 2007; Manski 2007; Van den Bulte 2010). While different labels have been used to capture the subtle differences among them, a key argument that the preferences and/or behaviors of a person may be influenced by those of others is shared.

Understanding the peer effects in consumer decisions is important for proposing policies that can improve the desired outcomes from the perspective of the marketer and policy maker because the policies that create and influence the contagion effect could yield benefits of expanding the total effect of policies through social multipliers (Glaeser, Sacerdote, and Scheinkman 2003).

* This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2014S1A5A8019666)

** Associate Professor of Marketing, Ewha School of Business, Ewha Womans University (songs@ewha.ac.kr)

In this research, we seek to find evidence for peer effects in consumers' health-related activities – specifically, gym-going behaviors. Would the workout frequency of other members at a gym have any influence on a focal member's workout behaviors? That is, how frequently one would work out at a gym may be a function of not only the individual's characteristics such as age and self-discipline but also the behaviors and characteristics of other members at the branch. While obesity had not been thought of as an infectious condition, Christakis and Fowler (2007) first suggested obesity might spread through social network due to peer influences.

Similarly, Ball and colleagues (2011) surveyed over 3,000 people and reported that those who have network ties with healthy people tend to exhibit better health-related behaviors in eating, exercising and dropping bad habits, which also suggests that healthy behavior may be contagious.

Our contribution to this stream of research is two-fold: first, we show that there are two kinds of peer effects that may be in action – behavioral and contextual peer effects. Second, instead of just capturing the peer effects at the mean, we capture the distributional impact of peer effects using quantile regression, which allows us to characterize the heterogeneous effects on different points of an outcome distribution.

Ⅱ. Peer Effects

Clustering of consumer behaviors over space and time has been reported in a number of studies but whether it is indeed caused by the peer influence has been called into question (Manski 1993; 2007; Moffit 2001; Brock and Durlauf 2001; Blume et al. 2010). Many economists have regarded social contagion to be spurious phenomena and argued that it is merely an artifact of confounding effects (Manski 1993; Blume et al. 2010). For example, Cohen-Cole and Fletcher (2008) were able to replicate the findings of Christakis and Fowler (2007), *i.e.*, contagion in obesity through the social network, but they also found implausible spurious social contagions in height, acne, and headache, using similar data and methodology.

Marketers and social scientists have been interested in examining opinion leaders' influence (Nair, Manchanda, and Bhatia 2010; Trusov, Bodapati and Bucklin 2010). For example, Nair and colleagues surveyed physicians to identify those who most influence other doctors' decisions, and showed marketing to opinion leaders can boost revenues by an average of 18 percent.

Few researchers in marketing, however, have successfully addressed homophily — endogenous group formation and simultaneity problem. Researchers have largely disregarded the seriousness of the problem and simply argued these problems should be less of a concern in their specific empirical setting in which social contagion is tested (*e.g.*, Du and Kamakura 2011, Bell and Song 2007). It is mainly because research on social contagion in marketing has mainly focused on the issue of new product adoption, which made marketing researchers rely on the data with limited variations within individuals, e.g., cross-sectional observations across individuals. That is, the dependent variable used in the model is either duration or dichotomous variable with a series of zeros for yet-to-adopt behavior and one for adoption. The lack of variation over time in the outcome measures makes it difficult for researchers to employ a fixed-effects at the level of individual. This lack of degrees of freedom simply makes impossible to tease the social interaction effects from other confounders unless strong assumptions about the process are imposed.

Another common data limitation is related to the need for exogenously-defined reference group in peer effect estimations. Researchers have defined reference group based on either geographically proximate location (e.g., the number of recent purchasers among the nearest neighbors in Du and Kamakura 2011; Bell and Song 2007), or social network directly elicited via surveys (*e.g.*, Nair et al. 2010; Igenyar et al. 2010). There are studies in which researchers turned to surveys to collect data on reference groups and/or ties between social network members. For example, Nair et al. (2010) collected surveys from physicians and opinion leaders asking them directly who they turn to for advice in their analysis of the spread of prescription drugs. While surveys allow researchers to capture ties in the social network, they pose a serious identification problem. When the reference group is not defined exogenously, however, separating endogenous group formation from causal network effects becomes difficult, and we cannot tell if the observed correlation is due to a social contagion or a manifestation of homophily.

In this paper, we address these identification issues in estimating peer effects. The richness of our data and the flexibility and robustness of our modeling approach allow us to rule out alternative explanations. Specifically, the panel structure of our data enables us to track how the gym usage changed over the observation periods. We use within-person variations in gym usage instead of the between-person variations in order to estimate peer effects, addressing the difficulties of identification due to homophily and simultaneity. By explicitly controlling for individual difference in workout frequency using individual-level fixed effects, we address problems of endogenous group formation and locally-targeted marketing activities. By including time-fixed effects for each monthly period, we hold constant cross-temporal variation which could confound their results. We also address the simultaneity issue by employing instrumental variables (IV) estimation.

Even with all these controls, however, we cannot completely rule out the possibility that there could be other sources of observed correlated patterns in members' behaviors. For instance, if there exist temporary marketing efforts at certain branches, omission of this branch-specific confounding may result in upward bias in the estimated parameters. We test robustness of our findings through falsification tests and various alternative model specifications.

Regarding the direction of behavioral peer effects, we believe there could be opposite forces making the effect both positive and negative. On one hand, the behavioral peer effects could be positive. Studies have shown that people look for other people to mimic the behavior of people around them and use it as a guide to their behavior (Goldstein, Cialdini, and Griskevicius 2008). Peer effects could operate below conscious level, without a person realizing being influenced. One the other hand, one might expect the peer effects in usage to be negative because there might negative network externalities due to the crowdedness at the branch. Because there is a fixed number of workout equipment, the more number of people work out at the branch, the less enjoyable the workout becomes (Desor 1972). Last but not least, we hypothesize that the degrees to which individuals be influenced by peers would be different among the members. Specifically, the impact of behavioral and contextual social influence would vary across members depending on his or her level of workout frequency.

Ⅲ. Data

The main body of our data comes from Crunch Gyms, one of the largest health club chains in the United States. We obtained individual-level data on weekly attendance, dates of enrollment, termination of the contract, contract type, branch of enrollment, gender, birth date (age), corporate subsidy, enrollment fee, and monthly dues for 86,434 members at 13 branch locations in New York City. Figure 1 shows the location of Crunch Gyms' thirteen branches in New York City. The observation period spans from July 2006 to May 2008 for the duration of 102 weeks, resulting in 3,474,066 person-week panel data. We also augmented neighborhood characteristics using zip-code information from 2000 US Census data to create additional control variables. These variables include population, the proportion of Black, Hispanic, single male households, single female households, those with bachelor's or higher degrees, households with fulltime male, households with full-time female, households with income greater than 75,000 dollars, homes valued at greater than 250,000 dollars, and homes with more than 5 bedrooms.

In Figure 2, we present the histogram of weekly visits per a gym member. It shows

〈Figure 1〉 Locations of Crunch Gym in NYC

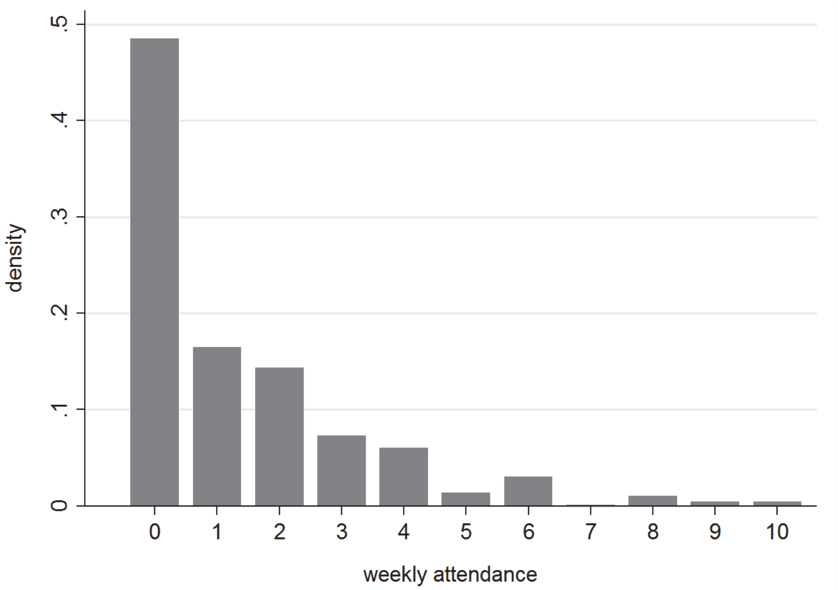
that roughly half the members do not work out at all in any given week. The average number of visits per member is 1.37. Table 1 presents the summary statistics of membership data. The average age of the members is 33.8, and 44 percent of the members are male. 14 percent of the members have the membership fee subsidized by their employers. There are significant variations across thirteen branch locations, which hints a possibility of members' self-selection into different branches. The changes in average weekly visits across the branches in 2007 are illustrated in Figure 3.

This data has many merits for the study of peer influence. First, the data provides a well-defined reference group and instrumental variables and allows us to calculate the workout frequency and characteristics of the peers who are physically

〈Table 1〉 Summary Statistics

branch	n	attend	duration	age	male	corp	defect	monthly dues	enrollment fee	annual	monthly commit	monthly flex
NY101	5,579	1.38	39.8	33.0	0.36	0.08	0.60	81.8	29.6	0.31	0.50	0.16
NY102	4,789	1.46	37.3	35.8	0.50	0.35	0.55	76.8	42.1	0.27	0.43	0.29
NY103	6,011	1.32	37.4	33.5	0.44	0.20	0.58	82.7	33.4	0.27	0.48	0.22
NY104	3,129	1.38	38.3	34.1	0.59	0.20	0.58	80.8	35.8	0.27	0.43	0.28
NY105	7,756	1.50	39.5	33.3	0.41	0.13	0.58	83.7	32.1	0.30	0.48	0.19
NY106	8,227	1.50	43.2	36.2	0.41	0.06	0.50	85.4	38.6	0.27	0.48	0.22
NY107	4,333	1.45	42.9	36.1	0.51	0.07	0.56	83.1	35.7	0.32	0.42	0.23
NY110	7,931	1.00	37.1	34.0	0.42	0.06	0.12	84.0	61.5	0.07	0.73	0.20
NY112	6,797	1.43	42.4	32.6	0.43	0.19	0.58	78.3	44.9	0.19	0.58	0.20
NY113	11,582	1.67	43.4	33.5	0.48	0.20	0.53	84.7	34.8	0.29	0.40	0.26
NY114	9,217	1.00	42.8	34.4	0.41	0.05	0.10	83.8	18.4	0.07	0.70	0.22
NY115	4,850	1.29	37.1	32.6	0.44	0.19	0.59	82.7	31.4	0.29	0.49	0.19
NY116	6,233	1.37	35.5	30.2	0.42	0.08	0.16	87.0	22.6	0.11	0.65	0.22
Total	86,434	1.37	40.2	33.8	0.44	0.14	0.44	83.3	35.4	0.23	0.53	0.22

〈Figure 2〉 Histogram of Weekly Attendances



co-located on a given week. Another important aspect of the data is longitudinal panel data, which allows us to employ fixed-effects models at the individual level. We also have information about the timing of enrollment which we make use of to measure the peer influence over time.

Ⅳ. Model

We estimate the linear-in-means model using random-effects and fixed-effects panel data regressions, accounting for possible correlated errors to reflect clustering at the gym branch level. Specifically, we use two-stage least-squares random-effects and fixed-effects (*i.e.*, within) estimator. Our model specification for a member i's workout frequency at the gym g and week t is as follows:

$$y_{igt} = \alpha_i + \delta_t + \beta x_i + \underbrace{\gamma \frac{1}{n_g - 1} \sum_{j \neq i}^{n_g} y_{jgt} I(j \in g)}_{\text{behavioral peer effects } (\gamma \bar{y}_{(-i)gt})} + \underbrace{\mu \frac{1}{n_g - 1} \sum_{j \neq i}^{n_g} x_{jgt} I(j \in g)}_{\text{contextual peer effects } (\mu \bar{x}_{(-i)gt})} + \epsilon_{igt}$$

where i indexes a gym member, g indexes a gym branch, and t indexes time. The α_i's are the time-invariant individual-specific effects, δ_t are the time fixed-effects, x_i is the individual-level and zip-code level demographics, and ε_{ijt} is the random error term.

The behavioral peer effects are represented as the average workout frequency at the member's branch $y_{(-i)gt}$ and the contextual peer effects as peer characteristics at the member's branch $x_{(-i)gt}$. The subscript $(-i)$ means that the average values are computed over all members in the branch excluding the member i. To mitigate the reflection problem due to the bi-directional influence between a gym member and the peers at the branch, we deliberately exclude member i's own workout frequency when we construct the branch average, preventing the mechanical incorporation of a members' own frequency from making into the mean.

4.1 Instrumental Variable Model

To account for the endogeneity problem of the behavioral peer effects, we use instrumental variable regression models. Instrumental variables (IV) regression is a general way to obtain a consistent estimator when the regressor, x, is correlated with the error term, ε. We need an instrument that has no effect on outcomes (*i.e.*, a member's attendance) other than through the first-stage channel. Therefore, we use average workout frequency at other branches as IVs: *IV*1, average attendance at all the other locations in NYC, and *IV*2, average attendance at the nearest location from the focal gym. It should be noted that peer effects occur within a branch, not across the branches (spurious contagion effects).

4.2 Alternative Models

The group mean outcome is a linear function of the group mean characteristics in the linear-in-means model, and any correlations between individuals' outcome and her reference groups' mean outcome may simply reflect the effects of the group mean characteristics. We test alternative nonlinear model specifications to check the robustness of our findings. First, we run a fixed-effects negative binomial (NBD) regression model to see whether our finding is driven by the linearity assumption on our empirical specification. This model also accounts for the count nature of our dependent variable, workout frequency. We also fit an instrumental variable (IV) Poisson regression model estimated with a two-step Generalized Method of Moments (GMM).

4.3 Distributional Effects of Peer Influences

Literature has mainly focused on estimating the peer effects at the mean but this approach may not capture the complete picture. While the conditional mean regressions provide a summary of the impact of the covariates, they fail to describe the full distributional impact on different parts of the distribution — unless the variable affects both the central and the tail quantiles in the same way. For example, Imbens and Agrist (1994) show that the wage distribution can become more compressed or the upper tail inequality may increase while the lower tail inequality decreases. The distribution of the dependent variable may change in ways that are not completely revealed by an examination of averages.

4.4 Quantile Regression

We use quantile regression to estimate the effects over the entire outcome distribution and to characterize the heterogeneous impacts of variables on different points of an outcome distribution (Koenker and Bassett 1978). It generalizes Laplace's median regression and can be used to measure the effect of covariates not only on the center of a distribution but also on the tails of the outcome distribution. It is more robust to distributional assumptions and can identify more subtle effects which would be missed by the application of mean regression. We model heterogeneity of effects via a quantile regression model and evaluate whether the peer effects are concentrated in the lower or upper end of the distribution.

Using the notation in the equation (1), τ_j-th conditional quantile function, $Q(\cdot|\cdot)$, is

$$Q_{y_{igt}}(\tau|\overline{y}_{-igt}, \overline{x}_{-igt}, \alpha_i, \delta_t) = \alpha_i(\tau) + \delta_t(\tau) + \beta(\tau)x_i + \gamma(\tau)\overline{y}_{-igt} + \mu(\tau)\overline{x}_{-igt}$$

where τ is a quantile in (0, 1). By estimating parameters $\gamma(\tau)$, $\mu(\tau)$, and $\beta(\tau)$, we can see how the two peer effect variables and other independent variables influence the location, scale, and shape of the conditional distribution of our workout frequency measure.

For estimation, we use fixed-effects estimator proposed by Koenker (2004). Let $C(\cdot)$ be the summation of quantile regression check functions.

$$C(\tau;\gamma,\mu,\alpha,\delta)=\sum_{i}^{N}\sum_{g}^{G}\sum_{t}^{T}\rho_\tau(y_{igt}-\alpha_i(\tau)$$
$$-\delta_t(\tau)-\beta(\tau)x_i-\gamma(\tau)\overline{y}_{-igt}-\mu(\tau)\overline{x}_{-igt})$$

where $\rho\tau$ is the classical quantile regression check function (see Koenker (2005) for more detail). We follow Koenker's approach for the joint estimation of the parameter of interest and the nuisance parameters, and the estimator is defined as,

$$(\hat{\gamma}(\tau),\hat{\mu}(\tau),\hat{\alpha}(\tau),\hat{\delta}(\tau))=\textbf{argmin}\ \ C(\tau;\gamma,\mu,\alpha,\delta).$$

To address potential endogeneity problem, we use the instrumental variable quantile regression model developed by Koenker (2004) and Harding and Lamarche (2009) — a panel data model with endogenous independent variables, where we allow endogenous variables to be correlated with unobserved factors affecting the response variable.

V. Results

5.1 Initial Evidence from OLS regression

Table 2 presents the results from the initial step of our analysis — the OLS estimation from equation (1), without and with zip-code level control variables. Both models show strong, positive behavioral peer effects (γ) and the estimated coefficient in model 1 is 1.009 and in model 2 is 0.998. Contextual peer effects (μ_{age} and μ_{gender}) are estimated to be significant, yet weaker and negative. The estimation results indicate that the average age of the members and the proportion of male members are negatively correlated with a focal member's workout frequency. Coefficients of all three individual characteristics (age, gender, and corporate subsidy status) are highly significant, and it suggests that those who are older, male, and with the subsidy are likely to workout more frequently. OLS estimates, however, could be biased because members' choice of a branch is not random. It is possible that members sort themselves into a particular location according to their preferences for certain gym equipment and atmosphere creating greater similarity amongst the members at a given gym branch, which then could generate spurious positive correlation. There is a reason to believe this might have been the case as great heterogeneity is observed across different locations in gym usage (see Figure 3).

5.2 Instrumental Variable Model

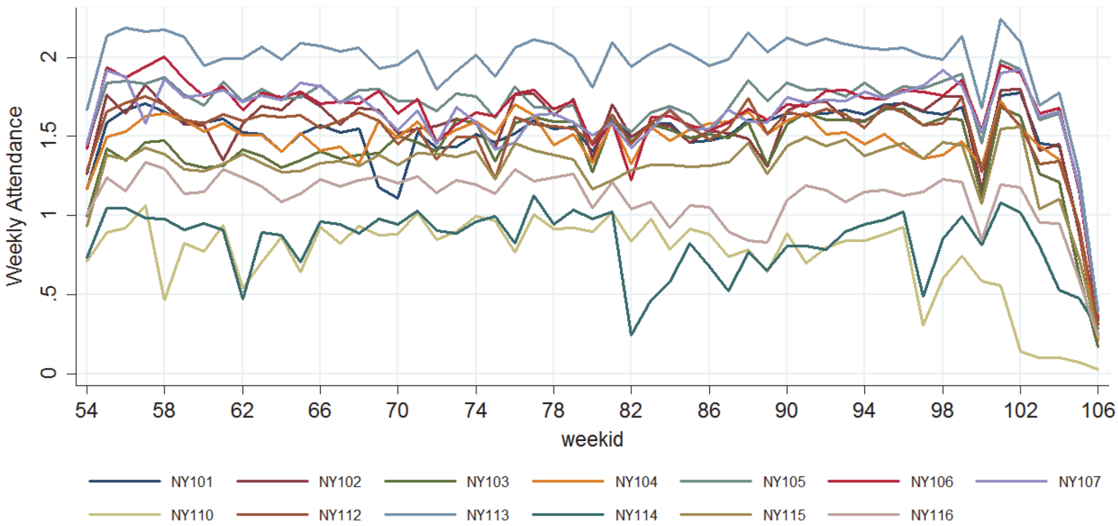
We use instrumental variables to address the potential biases from simultaneity. We use the

〈Table 2〉 OLS Estimation Results

	(1)				(2)			
	Coefficient	Std. Err.	t	$P > \|t\|$	Coefficient	Std. Err.	t	$P > \|t\|$
γ	1.009	0.003	311.482	0.000***	0.998	0.003	285.340	0.000***
μ_{age}	-0.004	0.000	-11.728	0.000***	-0.015	0.000	-34.396	0.000***
μ_{gender}	-0.255	0.025	-10.027	0.000***	-0.154	0.030	-5.129	0.000***
gender	0.164	0.003	59.529	0.000***	0.162	0.003	55.007	0.000***
age	0.022	0.000	161.394	0.000***	0.022	0.000	152.168	0.000***
corporate	0.214	0.004	53.888	0.000***	0.254	0.004	59.350	0.000***
% Black					-0.156	0.010	-15.635	0.000***
% Hispanic					-0.057	0.015	-3.701	0.000***
% single male HH					1.584	0.100	15.824	0.000***
% single female HH					0.135	0.071	1.903	0.057
% with bachelors or higher					0.001	0.040	0.023	0.982
% HH w/ fulltime male					-0.178	0.037	-4.824	0.000***
% HH w/ fulltime female					-0.442	0.041	-10.745	0.000***
% wealthy HH ($\geq$ \$75k)					0.250	0.049	5.099	0.000***
population					0.000	0.000	8.022	0.000***
% homes valued at $\geq$ 250k					-0.038	0.046	-0.838	0.402
% home with 5+ bedrooms					-1.291	0.149	-8.668	0.000***
intercepts	-0.615	0.013	-47.173	0.000***	-0.254	0.022	-11.498	0.000***
r^2	0.054				0.058			
N	3,077,572				2,729,037			

*** Significant at the 0.001 level, ** Significant at the 0.01 level, * Significant at the 0.05 level.
Robust standard errors are clustered by gym branch.

〈Figure 3〉 Average weekly attendance across branches in 2007



mean workout frequency by members at all the other branches in New York (*attend_allothers*) and by members of the nearest branch (*attend_nearest*). If the instruments are only weakly correlated with the endogenous explanatory variable, then even a weak correlation

〈Table 3〉 Correlation Table (Instrumental Variables)

	attendances	attend_ic	attend_allothers	attend_nearest
attendances	1.0000			
attend_ic	0.2199	1.0000		
attend_allothers	0.1145	0.5168	1.0000	
attend_nearest	0.1542	0.6963	0.6374	1.0000

〈Table 4〉 Panel IV models - First Stage Estimation Results

	First Stage Results			
dependent variable: attend_ic	Coefficient	Std. Err.	z	$P > \|z\|$
attendance (at all other branches)	.767	.001	1060.62	0.000***
attendance (at the nearest branch)	.206	.001	389.56	0.000***
μ_{age}	.023	.000	222.02	0.000***
μ_{gender}	.373	.005	69.51	0.000***
gender	.014	.001	12.70	0.000***
age	.001	.000	13.90	0.000***
corporate	.069	.002	43.27	0.000***
% Black	-.353	.004	-97.83	0.000***
% Hispanic	-.111	.006	-19.98	0.000***
% single male HH	-.418	.036	-11.68	0.000***
% single female HH	1.130	.026	43.59	0.000***
% with bachelors or higher	-.243	.014	-16.84	0.000***
% HH w/ fulltime male	-.213	.013	-15.81	0.000***
% HH w/ fulltime female	-.312	.015	-20.83	0.000***
% wealthy HH ($\geq$ \$75k)	.254	.018	14.26	0.000***
population	.000	.000	18.61	0.000***
% homes valued at $\geq$ 250k	.319	.016	20.12	0.000***
% home with 5+ bedrooms	-2.030	.051	-39.89	0.000***
timedummy1	-.092	.001	-125.29	0.000***
timedummy2	-.079	.001	-105.75	0.000***
timedummy3	-.060	.001	-77.64	0.000***
…				
timedummy24	-.001	.001	-1.15	0.250***
intercepts	-.621	.007	-88.94	0.000***
N		2,729,037		

*** Significant at the 0.001 level, ** Significant at the 0.01 level, * Significant at the 0.05 level.
Robust standard errors are clustered by gym branch.

between the instruments and the error in the original equation can lead to a large inconsistency in IV estimates. We first checked the correlation between the outcome variable (*attendances*), endogenous variable (*attend__ic*), and two instrumental variables (*attend__allothers* and *attend__nearest*). Table (3) shows that the two IVs are strongly correlated with the endogenous variable (0.5168 and 0.6963).

The results from the first-stage regression of endogenous variable (attend__ic, $y_{(-i)gt}$) on the excluded instruments are presented in Table 4. The F-statistics strongly reject the null that the exogenous instruments have no explanatory power. The strength of the (first-stage) excluded instruments is of critical importance in obtaining consistent estimates when 2SLS estimation is used (Stock, Wright, and Yogo 2002) because if instruments are weak, the coefficients are biased toward inconsistent OLS

〈Table 5〉 Panel IV models (RE vs. FE) Estimation Results

	Random Effects IV estimator				Fixed-effects IV estimator			
	Coefficient	Std. Err.	z	$P > \|z\|$	Coefficient	Std. Err.	z	$P > \|z\|$
γ	0.972	0.007	142.310	0.000***	0.973	0.007	140.386	0.000***
μ_{age}	-0.016	0.001	-12.791	0.000***	0.005	0.002	2.675	0.007**
μ_{gender}	-0.448	0.066	-6.817	0.000***	-0.461	0.074	-6.210	0.000***
gender	0.139	0.014	10.226	0.000***				
age	0.008	0.001	11.930	0.000***				
corporate	0.230	0.020	11.781	0.000***				
% Black	0.040	0.044	0.897	0.370				
% Hispanic	0.011	0.068	0.163	0.871				
% single male HH	0.825	0.440	1.874	0.061				
% single female HH	-0.130	0.321	-0.407	0.684				
% with bachelors or higher	0.284	0.178	1.596	0.111				
% HH w/ fulltime male	-0.062	0.166	-0.374	0.708				
% HH w/ fulltime female	-0.070	0.184	-0.383	0.702				
% wealthy HH ($\geq$ \$75k)	-0.211	0.220	-0.962	0.336				
population	0.000	0.000	0.340	0.734				
% homes valued at $\geq$ 250k	-0.206	0.195	-1.052	0.293				
% home with 5+ bedrooms	0.903	0.628	1.438	0.150				
timedummy1	0.359	0.009	39.373	0.000***	0.388	0.009	41.830	0.000***
timedummy2	0.502	0.010	52.834	0.000***	0.533	0.010	55.661	0.000***
timedummy3	0.543	0.010	55.707	0.000***	0.574	0.010	58.540	0.000***
…								
timedummy24	0.124	0.008	16.499	0.000***	0.138	0.008	18.129	0.000***
intercepts	-0.230	0.085	-2.706	0.007**	-0.287	0.069	-4.177	0.000***
N		2,729,037				2,729,037		

*** Significant at the 0.001 level, ** Significant at the 0.01 level, * Significant at the 0.05 level.
Robust standard errors are clustered by gym branch.

estimates. The results from the first-stage regression suggest that weak identification should not be a problem.

The results for the random- and fixed-effects IV regression appear in Table 5. Hausman test results indicate that the null hypothesis of difference in coefficients not systematic is rejected. The fixed-effects model results show that there exist strong behavioral peer effects (γ = 0.973, z = 140.39) and the contextual peer effects are present but weaker (μ_{age} = 0.005 and μ_{gender} = −0.461). Regarding the effects of other individual-level characteristics, we find that members who are male, older, and with corporate subsidies work out more frequently.

The results of random- and fixed-effects Negative Binomial model with panel data appear

〈Table 6〉 Negative Binomial Regression (RE vs. FE) Estimation Results

	Random Effects NBD model				Fixed-effects NBD model			
	Coefficient	Std. Err.	z	$P > \lvert z \rvert$	Coefficient	Std. Err.	z	$P > \lvert z \rvert$
γ	0.693	0.004	196.183	0.000***	0.701	0.004	194.125	0.000***
μ_{age}	-0.036	0.001	-62.004	0.000***	-0.044	0.001	-61.612	0.000***
μ_{gender}	-1.151	0.034	-33.467	0.000***	-1.263	0.038	-33.386	0.000***
gender	0.088	0.005	19.136	0.000***	0.078	0.006	13.459	0.000***
age	-0.001	0.000	-5.161	0.000***	-0.002	0.000	-9.156	0.000***
corporate	-0.182	0.006	-29.166	0.000***	-0.357	0.007	-49.036	0.000***
% Black	0.047	0.016	2.951	0.003**	0.144	0.021	6.868	0.000***
% Hispanic	0.284	0.024	11.684	0.000***	0.428	0.032	13.539	0.000***
% single male HH	1.336	0.153	8.750	0.000***	1.429	0.191	7.463	0.000***
% single female HH	-0.280	0.110	-2.556	0.011*	-0.332	0.138	-2.398	0.016*
% with bachelors or higher	-0.322	0.062	-5.193	0.000***	-0.525	0.080	-6.581	0.000***
% HH w/ fulltime male	0.534	0.056	9.469	0.000***	0.650	0.070	9.241	0.000***
% HH w/ fulltime female	0.224	0.065	3.463	0.001***	0.690	0.085	8.133	0.000***
% wealthy HH ($\geq$ \$75k)	0.356	0.075	4.736	0.000***	0.469	0.095	4.951	0.000***
population	0.000	0.000	1.235	0.217	0.000	0.000	2.606	0.009**
% homes valued at $\geq$ 250k	-0.410	0.070	-5.857	0.000***	-0.668	0.089	-7.472	0.000***
% home with 5+ bedrooms	0.643	0.231	2.777	0.005**	1.409	0.307	4.590	0.000***
timedummy1	0.335	0.005	71.746	0.000***	0.375	0.005	78.754	0.000***
timedummy2	0.446	0.005	83.919	0.000***	0.492	0.005	91.151	0.000***
timedummy3	0.472	0.006	85.264	0.000***	0.518	0.006	92.500	0.000***
…								
timedummy24	0.076	0.004	19.040	0.000***	0.082	0.004	20.275	0.000***
intercepts	0.851	0.032	26.660	0.000***	0.961	0.039	24.574	0.000***
N		2,729,037				2,721,024		
LL		-3990963.315				-3691593.846		

*** Significant at the 0.001 level, ** Significant at the 0.01 level, * Significant at the 0.05 level.
Robust standard errors are clustered by gym branch.

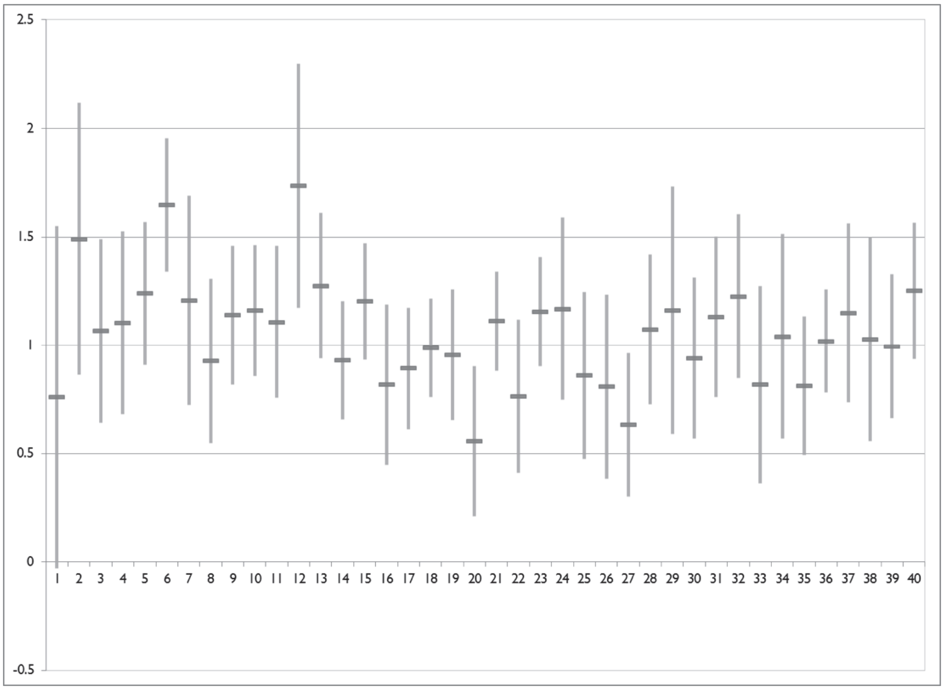
in Table 6 and IV Poisson Model in 7. The results parallel those from Table 5. The coefficient for behavioral peer effects shows positive sign and two contextual effects are estimated to be negative.

To further rule out the possibility of detecting spurious peer effects, we looked at the subset of membership data in which we only kept those who had just joined the gym. We expect the peer effects in the their first week to be minimal and statistically non-significant because these newly-joined members would have had little time and opportunities to observe the other members' behaviors at the branch. The results show that the peer effects are minimal right after a member joins the gym with the estimated coefficients not being statistically significant, which is consistent with our expectation (See Figure 4, 5, 6).

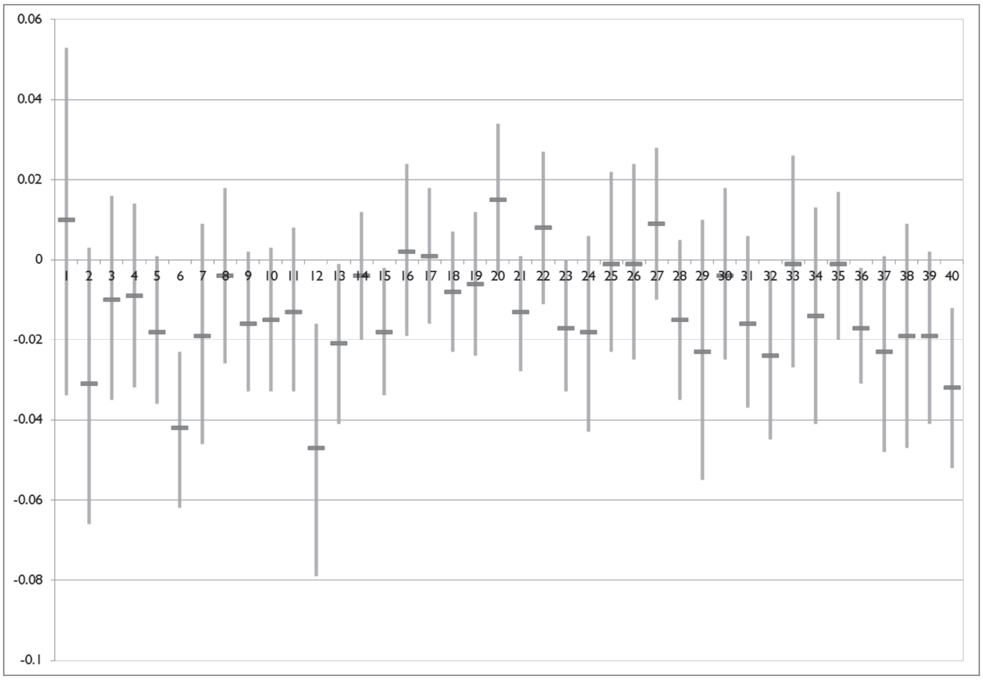
5.3 IV Quantile Regression

The estimated peer effects from the instrumental variable (IV) quantile regression are reported in Table 8 (Harding and Lamarche 2009). The results show that the behavioral peer effects are positive and significant at the 50th percentile and above (*i.e.*, in the upper tail of outcome distribution), but the peer influence is close to zero in the lower tail. They indicate that the

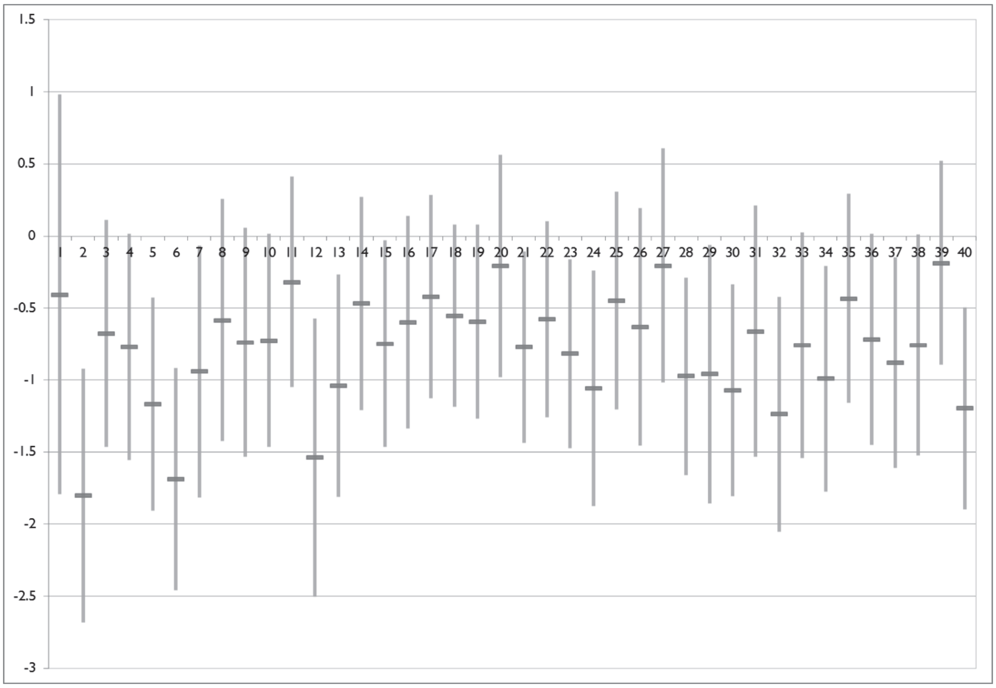
〈Figure 4〉 Behavioral peer effects for newly joined members (week 1 - 40)



〈Figure 5〉 Contextual peer effects (μ_{age}) for newly joined members (week 1 – 40)



〈Figure 6〉 Contextual peer effects (μ_{gender}) for newly joined members (week 1 – 40)



〈Table 7〉 IV Poisson Model Estimation Results (2-step GMM estimator)

	IV Poisson Model (2-step GMM estimator)			
	Coefficient	Std. Err.	z	$P > \|z\|$
γ	1.033	0.023	45.014	0.000***
μ_{age}	-0.025	0.002	-13.159	0.000***
μ_{gender}	-0.870	0.128	-6.802	0.000***
gender	0.098	0.012	8.514	0.000***
age	0.013	0.001	23.661	0.000***
corporate	0.145	0.016	9.100	0.000***
% Black	-0.065	0.038	-1.727	0.084
% Hispanic	0.016	0.058	0.283	0.778
% single male HH	2.437	0.407	5.993	0.000***
% single female HH	-0.658	0.301	-2.187	0.029*
% with bachelors or higher	-0.246	0.152	-1.612	0.107
% HH w/ fulltime male	0.361	0.156	2.308	0.021*
% HH w/ fulltime female	-0.545	0.172	-3.168	0.002**
% wealthy HH ($\geq$ 75k$)	0.305	0.215	1.421	0.155
population	0.000	0.000	0.578	0.564
% homes valued at $\geq 250k$	-0.004	0.163	-0.026	0.980
% home with 5+ bedrooms	-0.984	0.534	-1.842	0.066
timedummy1	0.142	0.011	13.068	0.000***
timedummy2	0.385	0.020	19.379	0.000***
timedummy3	0.445	0.023	19.743	0.000***
…				
timedummy24	0.115	0.006	17.775	0.000***
intercepts	-0.968	0.092	-10.564	0.000***
N		2,729,037		

*** Significant at the 0.001 level, ** Significant at the 0.01 level, * Significant at the 0.05 level.
Robust standard errors are clustered by gym branch.

peer effect had a relatively large impact on the workout frequency of the focal member, and this impact is increasing in the quantile index. The instrumental variable quantile regression model estimates exhibit considerable heterogeneity, ranging from 0 to 2.30. One possible explanation is that members with very low level of utilization of the gym may not have had enough opportunities to observe and be influenced by other members at the gym branch. Regarding the contextual peer effects, we find that the effects of gender composition are negative (*i.e.*, the more there are at the branch, the less likely for one to workout) but they are statistically significant only at 50th and 75th quantile.

〈Table 8〉 IV Quantile Regression Estimation Results

	tau25	tau50	tau75	tau90
γ	0.000	0.409***	1.406***	2.300***
	(0.003)	(0.016)	(0.290)	(0.060)
μ_{age}	-0.000	-0.003*	-0.027***	-0.022***
	(0.000)	(0.002)	(0.002)	(0.004)
μ_{gender}	0.000	-0.237*	-0.293*	0.075
	(0.012)	(0.103)	(0.126)	(0.259)
gender	-0.000	0.138***	0.233***	0.305***
	(0.001)	(0.009)	(0.013)	(0.024)
age	-0.000	0.009***	0.026***	0.052***
	(0.000)	(0.001)	(0.001)	(0.001)
corporate	0.000	0.132***	0.307***	0.720***
	(0.002)	(0.014)	(0.021)	(0.044)
% Black	-0.000	-0.120***	-0.136***	-0.262***
	(0.003)	(0.030)	(0.038)	(0.073)
% Hispanic	0.000	-0.046	-0.082	-0.340**
	(0.006)	(0.049)	(0.060)	(0.111)
% single male HH	0.000	-0.013	0.737	2.028*
	(0.039)	(0.378)	(0.444)	(0.907)
% single female HH	-0.000	0.896	0.531	0.763
	(0.029)	(0.249)***	(0.317)	(0.608)
% with bachelors or higher	-0.000	0.042	0.160	0.755*
	(0.015)	(0.139)	(0.168)	(0.326)
% HH w/ fulltime male	-0.000	-0.243*	-0.148	-0.666*
	(0.015)	(0.120)	(0.153)	(0.297)
% HH w/ fulltime female	0.000	-0.117	-0.666***	-1.342***
	(0.016)	(0.125)	(0.171)	(0.331)
% wealthy HH ($\geq$ \$75k)	0.000	0.459**	0.349	-0.029
population	(0.019)	(0.175)	(0.219)	(0.405)
	0.000	0.000***	0.000	0.000
% homes valued at $\geq$ 250k	(0.000)	(0.000)	(0.000)	(0.000)
	-0.000	-0.309*	-0.183	0.099
% home with 5+ bedrooms	(0.017)	(0.139)	(0.183)	(0.325)
	0.000	-1.101	-1.810***	-2.062*
intercepts	(0.055)	(0.380)	(0.484)	(0.997)
	-0.000	0.566***	0.163	-0.800***
	(0.010)	(0.085)	(0.099)	(0.187)

*** Significant at the 0.001 level. ** Significant at the 0.01 level. * Significant at the 0.05 level.

Ⅵ. Discussion

Peer-to-peer interaction amongst customers has received much attention from marketing researchers in recent years and it will certainly stay as one of the important topics in marketing because customers are now connected to each other in numerous new ways. It also supplies rich data to researchers. The availability of individual level data allows us to understand the role of peer effects in consumer preferences and decisions. There is a growing literature on peer effects and social contagion in marketing, adoption of new product being studied most extensively. We contribute this body of research by examining peer effects in the domain of service usage, and we found evidence for peer effects in service usage in the context of customers' gym going behaviors.

It is important for marketers and policy makers to distinguish peer influences from alternative mechanism such as homophily which leads to a clustering of the outcome that mimic social contagion because it is critical for estimating viral marketing effectiveness or managing peer effects in service usage. Using a disaggregate membership data from one of the largest health club chains, we showed how other members at the gym influence a focal member's workout frequency. After correcting for endogeneity and potential selection bias, we find strong positive peer effects in gym members' workout behavior. When other members exhibit higher frequency of workout at a gym branch, it resulted in an increase in workout by the focal gym members. The peer effect remains substantial and statistically significant even after we include individual-level fixed effects, time fixed effects, and a large set of covariates at the individual and zip-code level, providing strong evidence for the peer influence in service usage. We also find the peer effects to be different in their magnitude for infrequent and frequent gym goers. Our results demonstrate the critical importance of properly identifying the relevant peer group when estimating peer effects. Our models correct for the endogeneity of individual and peer outcomes and rule out common shocks as the mechanism driving the peer effects.

This research suggests that managers have more reasons to monitor members' workout frequency with care and to encourage members to work out more frequently as these management activities could create ripple effects. That is, if managers of a gym devise a strategy to make some of its members come and work out more frequently, it would also affect those who are not directly targeted and the other gym members would also be likely to show up more frequently as the peer effects act as the social multiplier. Managers should also keep it mind that the social multiplier is like a double-edged sword; it also means that even if only a fraction of members start to skip the gym it could create a spiral effect making other members less likely to show up.

While it is likely that peers are sources of motivation and learning, we do not have a full account of the mechanisms though which peers affect outcomes due to the limitation of the data. Understanding how the characteristics of behaviors and products enable and constrain peer influences would be a valuable inquiry for future research. For example, the use of natural experiments or lab experiments (Christakis and Fowler 2010; Horton et al. 2011) may further help to clarify the extent to which contagion exists in social systems.

〈Received April 6. 2017〉
〈Revised April 21. 2017〉
〈Accepted April 24. 2017〉

References

Ball, Kylie, Robert W Jeffery, Gavin Abbott, Sarah A McNaughton and David Crawford, "Is healthy behavior contagious: associations of social norms with physical activity and healthy eating," *International Journal of Behavioral Nutrition and Physical Activity*, (in press).

Bell, David R., and Sangyoung Song (2007), "Neighborhood Effects and Trial on the Internet: Evidence from Online Grocery Retailing," *Quantitative Marketing and Economics*, 5 (4), 361-400.

Brock WA, SN Durlauf (2001), "Interaction-based Models," in *Handbook of Econometrics*, edited J Heckman and E Leamer, North-Holland, Amsterdam, 3297-3380.

Blume, Lawrence E., William A. Brock, Steven N. Durlauf, and Yannis M. Ioannides (2010), "Identification of Social Interactions," in *Handbook of Social Economics*, edited by J. Benhabib, A. Bisin, and M. Jackson, Amsterdam: Elsevier, 853-964.

Christakis, Nicholas A. and James H. Fowler (2007), "The Spread of Obesity in a Large Social Network over 32 Years," *New England Journal of Medicine*, 357, 370-379.

Christakis, Nicholas A. and James H. Fowler (2010), "Social Network Sensors for Early Detection of Contagious Outbreaks," PLoS One 5(9), e12948.

Cohen-Cole, Ethan, and Jason M. Fletcher (2008), "Detecting Implausible Social Network Effects in Acne, Height, and Headaches: Longitudinal Analysis," *British Medical Journal*, 337, a2533.

Desor, J. A. (1972), "Toward a psychological theory of crowding," *Journal of Personality and Social Psychology*, 21 (1), 79-83.

Du, Rex and Wagner Kamakura (2011), "Measuring Contagion in the Diffusion of Consumer Packaged Goods," *Journal of Marketing Research*, February, 28-47.

Glaeser, E. L., B. L. Sacerdote, and J. A. Scheinkman (2003), "The Social Multiplier," *Journal of the European Economic Association*, 1, 345-353.

Goldstein, N. J., R. B. Cialdini, and V. Griskevicius (2008), "A room with a viewpoint: Using social norms to motivate environmental conservation in hotels," *Journal of Consumer Research*, 35, 472-482.

Harding, Matthew, and Carlos Lamarche (2009), "A Quantile Regression Approach for Estimating Panel Data Models Using Instrumental Variables," *Economics Letters*, 104(3), 133-135.

Hedley, Allison A., Cynthia L. Ogden, Clifford L. Johnson, Margaret D. Carroll, Lester R. Curtin, Katherine M. Flegal (2004), "Prevalence of Overweight and Obesity Among US Children, Adolescents, and Adults, 1999-2002," *Journal of the American Medical Association*.

Horton J. J., D. G. Rand, and R. J. Zeckhauser (2011), "The Online Laboratory: Conducting Experiments in a Real Labor Market," *Experimental Economics*, 14, 399-425.

Iyengar, Raghuram, Christophe Van den Bulte, and Thomas W. Valente (2011), "Opinion Leadership and Social Contagion in New Product Diffusion," *Marketing Science*, 30 (2), 195-212.

Koenker, R (2004), "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91 (1), 74-89.

Koenker, R, and G Bassett (1978), "Regression Quantiles," *Econometrica*, 46, 33-50.

Lyle, David S. (2007), "Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point," *Review of Economics and Statistics*, 89 (2), 289-299.

Manski, Charles F. (1993), "Identificatin of Endogeneous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60, 531-542.

Manski, Charles F. (2007), *Identification for Prediction and Decision*, Cambridge, MA, Harvard University Press.

Moffitt, Robert A. (2001), Policy Interventions, Low-Level Equilibria, and Social Interactions, in Social Dynamics, edited by Steven N. Durlauf and Peyton Young, Brookings Institution Press, Washington D.C.

Nair, Harikesh, Puneet Manchanda, and Tulikaa Bhatia (2010), "Asymmetric Social Interactions in Prescription Behavior: The Role of Opinion Leaders," *Journal of Marketing Research*, 47 (5), 883-895.

Swinburn, Boyd, Gary Sacks, and Eric Ravussin (2009), "Increased food energy supply is more than sufficient to explain the US epidemic of obesity," *The American Journal of Clinical Nutrition*, 90 (6), 1453-1456.

Trusov, M., A. Bodapati, and R. E. Bucklin, (2010), "Determining influential users in internet social networks," *Journal of Marketing Research*, 47, 643-658.

Van den Bulte, Christophe, (2010), "Opportunities and challenges in studying customer networks," *In The Connected Customer: The Changing Nature of Consumer and Business Markets*, S. Wuyts, M.G. Dekimpe, E. Gijsbrechts, and R. Pieters, eds. Routledge, London, U.K., 7-35.