An Empirical Analysis of Individual Level Casino Gambling Behavior

Sridhar Narayanan
Puneet Manchanda*

July 2006
This version: June 2011
conditionally accepted, Quantitative Marketing and Economics

*Sridhar Narayanan is an Associate Professor of Marketing at the Graduate School of Business, Stanford University and Puneet Manchanda is the Isadore and Leon Winkelman Professor of Marketing at the Ross School of Business, University of Michigan. The authors are grateful to Wes Hartmann, Harikesh Nair, Subrata Sen, Katherine Burson, Scott Rick, seminar participants at Dartmouth's Tuck School, Duke University, London Business School, National University of Singapore, Rotterdam School of Management, Texas A&M University, University of British Columbia, University of Guelph, University of Iowa, University of Maryland, University of Michigan, University of Notre Dame, University of Texas at Austin, University of Toronto, University of Wisconsin at Madison and Yale School of Management, and the participants of the 2006 Summer Institute of Competitive Strategy and the 2009 Marketing Science Conference for comments and feedback and an anonymous company for providing the data and institutional insights. Manchanda would like to acknowledge the Kilts Center for Marketing and the True North Faculty Fund at the University of Chicago, and Narayanan would like to acknowledge the Coulter Family fellowship for providing research support. All correspondence may be addressed to the authors via e-mail at *sridhar.narayanan@gsb.stanford.edu* (Narayanan) or *pmanchan@umich.edu* (Manchanda).

Abstract

Gambling and gaming is a very large industry in the United States with about one-third of all adults participating in it on a regular basis. Using novel and unique behavioral data from a panel of casino gamblers, this paper investigates three aspects of consumer behavior in this domain. The first is that consumers are addicted to gambling, the second that they act on "irrational" beliefs, and the third that they are influenced by marketing activity that attempts to influence their gambling behavior.

We use the interrelated consumer decisions to play (gamble) and the amount bet in a casino setting to focus on addiction using the standard economic definition of addiction. We test for two irrational behaviors, the "gambler's fallacy" and the "hot hand myth" – our research represents the first test for these behaviors using disaggregate data in a real (as opposed to a laboratory) setting. Finally, we look at the effect of marketing instruments on the both the decision to play and the amount bet.

Using hierarchical Bayesian methods to pin down individual-level parameters, we find that about 8% of the consumers in our sample can be classified as addicted. We find support in our data for the gambler's fallacy, but not for the hot hand myth. We find that marketing instruments positively affect gambling behavior, and that consumers who are more addicted are also affected by marketing to a greater extent. Specifically, the long-run marketing response is about twice as high for the more addicted consumers.

Keywords: *Gambling Behavior, Addiction, Selection Models, Casino Gaming and Gambling, Slots, Hierarchical Bayes Methods*

1 Introduction

Gambling and gaming has evolved to becoming a very large and pervasive industry in the Unites States over the last three decades. In terms of size, it is estimated to have a total revenue of $ 73 billion in 2003. In terms of participation, a 2004 Gallup Poll showed that about half of all adult Americans bought a state lottery ticket and about one-third of all adult Americans visited a casino in 2003. The size and pervasiveness of the industry has led to an extensive debate about the benefits and costs of this industry to society. Supporters of the industry point to the fact that it provides an entertainment outlet for consumers, facilitates job creation, provides additional tax revenues and increases the welfare of disadvantaged groups in society such as Native Americans. Opponents of the industry argue that the rapid expansion of the industry creates opportunities that lead to undesirable behavior modification such as "addiction." Critics also assert that the industry takes advantage of "irrational" behaviors exhibited by consumers when they gamble. Finally, the industry is also criticized for its actions in providing incentives to consumers to facilitate gambling.

In this paper, using a rich, micro-level panel dataset of consumer betting activity at a casino, we investigate the above three charges frequently leveled against the gambling industry. Specifically, we endeavor to answer three questions. First, do consumers display addictive behaviors when it comes to gambling? Second, can we detect any evidence for irrational behavior on the part of the consumers when they gamble? Finally, is there any interaction between marketing activity and gambling behavior?

In order to test for patterns of addiction, we need to be precise about our definition of addiction. This is a somewhat challenging task as different literatures have proposed different definitions of addiction ranging from those based on physiological sources to propensities based on psychological drivers to purely behavioral outcomes. Given our data and setting, we choose to use a definition based on observed behavior that is consistent with the well accepted definition of addiction from the economics literature (Pollak 1970, Becker and Murphy 1988). The main idea behind this definition is intuitively appealing - addiction is deemed to be present when current consumption behavior is affected in a particular manner by past consumption behavior. Specifically, this literature defines addiction as the positive effect of past consumption on the marginal utility of current consumption. This dependence can arise when consumers are myopic (Pollak 1970) or forward-looking (Stigler and Becker 1977, Becker and Murphy 1988). The latter setting is referred to as "rational addiction," where consumers are forward-looking and therefore account for the effect of current consumption on future utility when making their consumption decisions. Using this definition, addictive behavior has been studied in a variety of contexts - cigarette smoking (Chaloupka 1991, Becker, Grossman, and Murphy 1994, Gruber and Koszegi 2001), caffeine consumption (Olekalns and Bardsley 1996) and gambling in state lotteries (Guryan and Kearney 2008).

The reduced form implication of the above definition is that there is a causal positive dependence in consumption across time periods (we refer the interested reader to Becker and Murphy (1988) or Guryan and Kearney (2005) for a detailed discussion on how the economic definition of addiction leads to this reduced form relationship).[1] In our context - casino gambling - we expect to find this positive effect of the volume of play (i.e. amount bet) in the past on a consumer's decision to play and bet an amount. Thus, controlling for other factors, for an addicted gambler, we expect to find that the consumer's likelihood of continuing to play and the amount bet is *increasing* in the amount of the previous bet. It is important to note that our chosen definition of addiction is by no means an exclusive definition. Our approach in this paper is to take as given this one well accepted definition and then to check in our data as to whether is evidence supporting this definition.

Besides addiction, we also focus on the possibility that consumers' observed gambles are consistent with irrational beliefs. Two beliefs that have been documented in the (experimental) literature are the "gambler's fallacy" and the "hot hand myth." The former refers to a mistaken belief in a negative correlation between two consecutive outcomes that are actually independent of each other. In contrast, the latter refers to the belief that attributes a series of positively correlated outcomes to some

[1]At this point, it should be clarified that we do not make a distinction between habit formation and addiction, both of which are forms of state dependence. Indeed, the literature on addiction itself has often used these terms somewhat interchangeably (Pollak 1970). Further, studies both in economics (Heckman 1981) and in marketing (Roy, Chintagunta, and Haldar 1996) have used habit persistence to refer to dependence of current choice on previous *propensity* to be in a state rather than the actual state itself and have distinguished this from structural state dependence, which is the dependence on the previous state. In our context, what we refer to as addiction is the latter effect. Overall, we are agnostic about whether this effect is termed addiction or habit formation and will use the term addiction henceforth to describe it.

underlying systematic (non-random) cause. Both beliefs are essentially based on people mistakenly believing in the representativeness of small samples. This underlying cause has been labeled the "law of small numbers" (Tversky and Kahneman 1971).

In our context, the gambler's fallacy implies that consumers are less likely to bet money if they have just won money and vice versa. The hot hand myth, in contrast, implies that the propensity to keep betting is higher the more the consumer has just won. It is worth noting that both these beliefs have been most often documented in laboratory experiments (e.g., Mullainathan (2002) and Gilovich, Vallone, and Tversky (1985)). While there is some evidence for this in the field, those studies have used aggregate data and are in different contexts than ours – state lotteries (e.g., Clotfelter and Cook (1993), Terrell (1994), Guryan and Kearney (2008)) and horse racing (e.g., Metzger (1984)).

Finally, to the best of our knowledge, the effects of marketing on gambling behavior at the individual level have not been studied in the literature. In particular, we examine whether in-casino marketing activity affects the gambling behavior of the consumer. Our individual-level analysis also allows us to examine correlations between addictive behaviors and the degree to which marketing affects gambling behaviors.

Our modeling approach accounts for two decisions made by a consumer on a trip to a casino – the decision to keep playing and the amount of the bet given that she decides to keep playing. An important feature of our model is that it is specified at the individual level in a hierarchical Bayesian framework. The advantage of using this approach is that we are able to obtain individual level parameters. This is important not only because it is novel in the context of casino gambling but because past research has suggested that the proportion of addicted gamblers is likely to be small. In addition, the two irrational beliefs referred to earlier, the gambler's fallacy and the hot hand myth, suggest seemingly contradictory patterns of behavior in the data and thus individual level analysis is likely to play a crucial role in determining the extent to which each of the beliefs is supported in the data.

Our main findings are as follows. We find that there is a small subset of gamblers (about 8% of the sample) who display evidence of addictive behaviors. Those classified as addicted don't differ significantly on demographic characteristics from the average gambler, but tend to be somewhat heavier gamblers on average. Marketing by the casino also tends to affect them to a greater extent. Furthermore, the long-run impact of marketing is significantly higher for these gamblers, both due to a higher direct effect of comps, and an additional indirect effect due to the effect incremental gambling has on future behaviors for these gamblers. An important finding in this paper is that we find evidence for the gambler's fallacy for a significant proportion of consumers. This, to our knowledge, is the first demonstration of this effect using individual-level data in a real-world gambling context. Finally, we find that a majority of consumers display behaviors that are consistent with gambling as an entertainment outlet, cashing out when the going is good and cutting their losses when it is not. We also find evidence for loss aversion in such behaviors, which is consistent with the predictions of Prospect Theory.

The rest of the paper is organized as follows. First, we describe the institutional features of the (casino) gambling industry and our data in section 2. Next, we discuss our modeling approach, the econometric specification and estimation in section 3. We discuss the results in section 4. We discuss the interaction between marketing and addiction in section 5. Finally, we carry out a robustness check in section 6 and conclude in section 7.

2 Industry Background and Data

As mentioned earlier, the gambling industry is estimated to have a total revenue of $ 73 billion in 2003. The industry as a whole can be divided into three major sub-categories. The largest sub-category is casino gaming accounting for 54% of this revenue. The next largest sub-category is that of state lotteries (25%). The balance is accounted for by a variety of smaller activities such as legal bookmaking and charitable games. In terms of consumer behavior, about half of all adult Americans buy a lottery ticket and about one-third of all adult Americans visit a casino every year (Gallup Poll 2003).[2] These numbers are confirmed by surveys conducted by casino operators e.g., a study carried out by Harrah's finds that one-fourth of all adult Americans has visited a casino in the last one year (Jones 2004). They also find that the average visit frequency is 5.8 times a year among gamblers. Interestingly, the anecdotal belief in the industry is that consumer participation has "saturated." This is confirmed by

[2]The details of this poll are referred to in Kearney (2005).

the same Harrah's study that the number of trips to casinos in the aggregate has grown at a negligible pace over the last four years. This has resulted in many casino companies trying to increase "share-of-wallet" and has increased the attention given to the role of marketing in the industry.

Our data come from a leading casino company operating primarily in the United States. The company operates properties that are spread nationwide (i.e., not just concentrated in Las Vegas). The company operates a loyalty card program to facilitate the building of relationships with its consumers. Enrollment in the loyalty card program is free and most consumers sign up for the card as it allows them to receive offers and benefits of membership. Membership in the loyalty program is ubiquitous – the company estimates that cardholders account for about 80% of all play in its properties. Typically, the consumer swipes the card before s/he begins play. All activity between then and the time the consumer exits the station is then uniquely linked to that customer's account and denoted as a "play."

The dataset consists of transaction activity for one property owned by the company.[3] This property is located in the southwestern United States. The property is a "local monopoly" as it is located about 30 miles from the nearest competitor casino (about 45 minutes driving time) and at an average distance of about 60 miles from other casinos in that part of the state. The property is open for operation twenty-four hours a day and seven days a week and offers play via slot machines, video poker, keno, bingo, blackjack and craps. The majority of amount bet - 90% - and a majority of transactions - 93% - in this property comes from slots. Given the nature of this property (it is not a "destination" casino), consumers make short and frequent trips to gamble at this property. The vast majority of these trips last a single day. In our subsequent discussion, we use the word "trip" to define a set of contiguous plays (over either a single day or a sequence of contiguous days) for an individual consumer.

The company markets its properties and games to consumers via coupons (redeemable for free hotel stays and/or meals) or cash equivalents. Consistent with industry practice, the company divides the use of these instruments into two broad classes. The first class is referred to as "offers." Offers essentially consist of all marketing activity that is aimed at influencing the decision to visit a given property (or

[3]Due to confidentiality concerns, we cannot reveal the name of the company and the geographic location of the property.

set of properties). The second class is referred to as "comps." Comps are geared towards increasing the duration of play and the amount bet once a consumer has already arrived at the property and has begun play. Thus, offers are typically given before the commencement of the trip, while comps are typically given during the trip. The company records the dollar value of comps delivered to each customer on each trip. Note that, in general, marketing dollars are not targeted randomly. The casino uses differences in past (total) activity across gamblers to decide its allocation of marketing dollars. As the past activity level increases for a given customer, the dollars spent also increase (usually in a step function).

Customer play data is available to us for a two-year period (Jan 2004-Dec 2005). The play data comprise all bet amounts by each consumer on each visit made to the casino. Besides the amount bet, we also observe the win/loss on each play, the comp amount delivered to the customer on a daily basis and some demographics of each consumer (age, gender, ethnicity/race).

The total number of consumers in our data is 198,223. These consumers comprise the universe of all players with loyalty cards who have played in this property during this time period. For the purpose of our analysis, we take a random sample of 2000 consumers from this set.[4] The total number of trips made by these consumers was 15632 with the number of mean trips per consumer being 7.8 (with a standard deviation of 17). This averages out to almost 4 trips per year which is close to the national average (5.8) described above.[5] In a trip, these consumers play on average 3.5 times (standard deviation 2.1) resulting in 56142 bets in total.

We provide descriptive statistics for trip level activity in Table 1. As can be seen from the table, the median trip lasts two hours, the total amount bet during a trip is \$ 450.00 (3.5 bets times \$ 131.00). There seem to be no "end-of-trip" effects on the amount bet - the mean amount bet in the last bet in trip is not significantly different from the mean amount bet for all other bets in the trip. The average hold of the casino is 16.54% (i.e. the casino keeps about 16.54 cents on average for every dollar bet). For our purpose, the relevant marketing activity are comps - the median comp

[4] A comparison of the demographic variables for the estimation sample, the full set of consumers in this casino and for all casinos owned by this firm shows that our estimation sample is representative of the gambling population for this casino chain.

[5] Note that the national average is across all casinos and is therefore expected to be higher than the number of trips to a particular casino, which is the case in our application.

value amount is 0. The mean comp amount is $ 9 and is 2% of the amount bet. Given that the casino hold is 16.5%, this suggests that the firm spends about 12% of its gross revenues on comps.

Table 1 about here

In terms of the demographics, the proportion of males is 45.9%, the mean age is 61 (standard deviation 13) and the racial/ethnic composition is as follows - Western European (64.6%), Hispanic (10%), Mediterranean (4.6%), Eastern European (3.7%) and Other (17.1%). Thus, the typical consumer in our sample is an older, white female. Another customer characteristic that we use in our analysis is a self-reported measure of whether the customer prefers to bet primarily at slot machines or at table games. 86% of consumers report that they prefer some form of slots, while 14% report that they prefer a table game. For our estimation sample, 91% of plays are on slot machines, constituting 89% of the total amount of bets placed.

3 Model

3.1 Model Development

We now discuss our model development. At the outset, we need to decide on the level of aggregation at which to carry out our analysis. Our approach is to conduct our analysis at the lowest level of aggregation available to us in a manner that we can best isolate the effects of past consumption on current consumption. As described earlier, the lowest level of data aggregation available to us is a play. A play starts when a consumer starts gambling at a station and ends when she exits the station.[6] Further, we restrict our attention to plays *within* a trip (recall that a trip is defined as the set of contiguous plays over either a single day or a sequence of contiguous days). What we mean by this is that while we use all observations across all trips for a consumer for estimation, we reset all variables (such as lagged and cumulative terms) at the beginning of the trip and do not model what happens between two trips.

[6] A potentially lower level of aggregation might be the sequence of activities within a play (e.g., each time a consumer puts a coin into the slot machine or places a bet at a game table) but the data are not retained by the firm at this level.

There are two reasons to conduct the within-trip play-level analysis as opposed to an across-trip level analysis. First, our data are from one particular casino. Consumers may visit another casino in the period between two trips to the casino in our dataset. Thus, activity at other casinos in the inter-trip period, which is not observed by us, could bias the estimated addiction effects. Second, since the mean number of trips per consumer is relatively small at 7.8 trips for the two-year period of the data, while there are more than 28 plays per consumer on average, the within-trip play-level analysis gives us greater data variation to reasonably conduct an individual-level analysis for a representative sample of our consumer. Note that as part of a robustness check, we also present the results from an across-trip analysis in Section 6.

Next we flesh out the three key aspects of our modeling approach for the play-level analysis - addiction, heterogeneity and selection. As mentioned earlier, we adopt the commonly accepted reduced form specification for addiction i.e., we allow for dependence in play and consumption across plays (bets) for a consumer within a trip to the casino. In our current formulation, we assume that consumers are myopic. In other words, current consumption is affected only by (immediate) past consumption. This is in contrast to models of rational addiction in which all consumption - past and expected future consumption - has an effect on current consumption. In these models, the general approach is to use instrumental variables for past and future consumption, since the rational expectation assumption makes future consumption endogenous. Such instruments are hard to find in our context. Further, the decision of whether to play again within a trip cannot be assumed to be exogenous (see the discussion below). These factors, coupled with the general computational difficulty of estimating individual level models with forward-looking behavior, lead us to estimate a model that conditions current consumption on only immediate past consumption.

One point to note here is that there could be inter-dependence between two adjacent bets due to inter-temporal substitution. Inter-temporal substitution comes about due to satiation in the consumer's utility function. If the consumer has been satiated in one consumption occasion, it may cause him to want to consume less on the next consumption occasion since he is still satiated. This is of particular importance in our context, where two consumption occasion (two plays) are very close to each other in time. Inter-temporal substitution would cause negative state dependence in consecutive plays, i.e. if the consumer has bet a high amount in the previous play, he is likely to bet less in the next one, or even stop playing. Thus, a negative coefficient on the parameter for the previous bet would be indicative of inter-temporal substitution. We cannot, given our data, separate between addiction and inter-temporal substitution and hence what we can find is the net effect of the two phenomena.

When testing for addiction, which manifests itself as a positive relationship between past and current bets, it is important to control for the amount of money that the gambler currently has available for betting. This amount of money (akin to a budget constraint) can potentially induce a negative correlation between successive bets. In order to eliminate a potential correlation arising from the budget constraint, we need to control for the amount of money that a consumer has available for a bet. While the wealth of the consumer is unobserved, we control for the amount of money in a consumer's pocket by introducing the total amount won or lost during the trip until the focal bet as a covariate in the model.

We allow for the effect of past wins and losses to be different, primarily to allow for loss aversion. Prospect theory (Kahneman and Tversky 1979) suggests that consumers react asymmetrically to wins and losses - in particular, their marginal utility for wins is lower than the marginal disutility for comparable losses. By interacting the amount of wins/losses with a dummy variable representing a win or a loss, we have different coefficients for wins and losses. This allows us to test for loss aversion empirically. We believe this is the first study to do this in the context of casino gambling using data outside the laboratory setting.

Another important factor in modeling betting behavior is that consumers' decisions to continue playing or exiting the casino is not just a function of previous betting behavior, but may be driven by fatigue and/or a time constraint. We control for these factors by including the time spent in the casino until the bet as a covariate. As mentioned earlier, a crucial aspect of our modeling approach is that we are able to estimate parameters at the individual level. We cast our model in a hierarchical Bayesian framework allowing us to obtain the posterior distributions of population as well as individual level parameters. This is crucial for both the detection of addiction as well as the irrational beliefs. For addiction, as previous research has found that only a small proportion of consumers behave in an addictive manner while gambling (Potenza, Kosten, and Rounsaville 2001), any model that estimates pooled parameters is likely to not find evidence for addiction. Furthermore, an analysis across individuals such as the one in Guryan and Kearney (2008) could potentially confound a local advertising/awareness spreading effect of a lottery win[7] with a hot-hand effect. Our focus on *within individual* effects eliminates this confound.

The specification of the model at the individual level also helps us control for the non-random nature of comps. As described earlier, the casino sets comps on the basis of past activity. In our model, this would be reflected in the correlation between two explanatory variables - the intercept and the comps variable. In our specification, the intercept is akin to an individual-level fixed effect,[8] thus soaking up all time-invariant cross-sectional differences. In other words, all information that is unique to the gambler (including the past level of activity) is captured by the fixed effect. Thus, the non-random nature of comps is reflected in the correlation between the individual level fixed effect and the comps delivered to that individual.[9] This correlation between two explanatory variables does not cause a bias in our estimates.

In terms of the irrational beliefs, while previous studies have documented the existence of these beliefs using behavioral data - e.g., the gambler's fallacy in Clotfelter and Cook (1993) and the hot hand myth in Guryan and Kearney (2008) - they have used aggregate data (on state lotteries). Typically, they have focused on these effects "across" agents. The disadvantage of such an approach is that there are alternative explanations such as local advertising/awareness effects that can result in similar outcomes, particularly in the case of the hot hand myth. By contrast, our specification of the individual level model allows us to carry out a "within" agent analysis that does not suffer from these potentially confounding effects. In addition, the individual level model can also help us distinguish between individuals (in the same sample) whose

[7]A jackpot win by a consumer in a local store, which is the context of Guryan and Kearney (2008), could lead to a greater awareness about that lottery in the local neighborhood through reports in local media, advertising by the lottery retailer, word of mouth etc. This could explain a positive effect of a jackpot win on subsequent purchases of lottery tickets at the store even in the absence of a lucky store / hot hand effect.

[8]As we describe in the results section, all the individual level intercepts, or fixed effects, are estimated very tightly. This therefore provides empirical support for this control as well.

[9]The casino chain classifies consumers into loyalty program tiers, based on their aggregate gambling volume across properties and over time. The comp plan for each individual is based on the tier that she belongs to. Note that our fixed effect arguments holds if the gambling volume for an individual consumer does not change over time in a manner such that comp plan for that consumer also changes. We were able to obtain this information for a set of representative gamblers at the casino property from which our estimation sample is drawn. We find that virtually all gamblers - 94.3% - remain in the same tier over the duration (two years) of our data.

behavior is consistent with one belief as opposed to the other.

Note that while our modeling approach allows us to find evidence for either belief, the literature pointing to the source of these biases suggests that, in our context, it is much more likely that we will find evidence for the gambler's fallacy. This is because the gambler's fallacy arises from a belief that small samples are representative of large samples and therefore, an outcome is more likely if it has not been observed in the past (or vice versa). In contrast, the hot hand myth, while also arising from the law of small numbers, originates from a slightly different consequence of the representativeness bias. This is the belief that if a sequence of positive outcomes is observed, it must be the case that the sequence is not random. In general, the gambler's fallacy likely applies to outcomes that agents believe to be generated from a random process, while the hot hand myth more likely applies to situations where agents can potentially attribute outcomes to something other than chance, for instance skill, motivation, energy levels etc. Experimental research has found that while agents believe in the gambler's fallacy for Roulette outcomes, they may simultaneously maintain a hot hand myth about their own predictions of the outcomes (Ayton and Fischer 2004). In our case, given that most of the plays are from slots ("games of chance") rather than table games ("games of skill"), it is more likely that we will find evidence for the gambler's fallacy rather than the hot hand myth.[10]

Finally, our model considers two decisions by the consumer at the end of each play - to continue playing and if so, the amount to bet. The consumer's decision of whether to continue playing and the amount that s/he bets cannot, in general, be considered independent. This leads to a classic selection problem. Specifically, if there are unobserved factors that affect both the decision whether to participate in a bet as well as the amount of the bet, the estimates from a regression of the bet amount would be biased. Hence, we need to control for this selection bias. Heckman (1979) proposed a two-stage estimator for correcting for selection bias. Given our Bayesian model, we cannot directly use this two-stage approach. Hence, we develop an approach where we directly model the correlation between the bet amount and

[10]Interestingly, using physical observations of players at a Roulette table for a few hours, Croson and Sundali (2005) document the presence of both the gambler's fallacy and the hot hand myth. While the evidence for the gambler's fallacy is quite clear in this study, that for the hot hand myth is less clear as an alternative account of a positive wealth effect of a past win on future bets is not controlled for in the study.

participation decisions. Thus, our model is a joint model of a consumer's decision to play (given that s/he has already played) and the amount bet. One point to note is that our analysis is conditional on the consumer choosing to visit the casino and play at least once during the trip.

In conclusion, our modeling approach builds upon previous work in the literature. Specifically, we use a panel data set of gamblers to obtain individual level parameters that affect their betting and continuation decisions within a trip. The distinctive features of our approach are that our model allows a reduced form test for addiction while controlling for many alternative explanations including irrational beliefs, marketing actions, cumulative outcomes (winnings/losses) and observable and unobservable differences across consumers (details below).

3.2 Model Specification

The model for the play-level analysis consists of two sub-models, one sub-model for the bet amount and another for the decision of whether to participate in the bet. We shall henceforth refer to the two as the *bet* sub-model and the *play* sub-model respectively. In general, our model specification consists of four sets of factors that could affect the decision to keep playing and the amount bet - the amount bet in the last bet, a win/loss in the last bet (with diminishing returns), cumulative winnings and losses up to the last bet (with diminishing returns) and the time spent in the casino. Additionally, we control for the effect of marketing activity by the casino - comps - in both the bet and play sub-models. As mentioned before, we interact the wins and losses in the previous bet with a dummy variable for whether it was a win. We also have interacted the cumulative wins or losses with a dummy variable for whether the cumulative amount is positive or negative. This specification therefore allows for loss aversion in the model.

The bet sub-model is given as

$$\ln(Bet_{it}) = X_{it}\alpha_i + \varepsilon_{it}, \text{ if } Play_{it} = 1 \tag{1}$$

where Bet_{it} is the amount of the bet in dollars for consumer i at play occasion t, $Play_{it}$ is an indicator variable for whether the consumer decides to participate in the bet or not, X_{it} is a vector of observable covariates, α_i is an individual-specific parameter vector and ε_{it} is an unobservable factor that affects the bet amount.

The covariates for the bet sub-model (X_{it}) include, besides an intercept, the bet amount in the previous bet within the trip, linear and quadratic terms for the amount won or lost in the previous bet, linear and quadratic terms for the cumulative amount won or lost in the trip until time period t and the amount of comps the consumer has received.

The definition of addiction we have adopted suggests that the coefficient for previous bet is a test for addiction. In particular, a positive coefficient would suggest that the individual displays evidence of addictive behaviors. The coefficients on the amount won or lost in the previous bet test for the hot hand myth or gambler's fallacy. In particular, if there is a positive coefficient of previous wins and/or a negative coefficient on losses, it suggests a belief in the hot-hand myth. A negative coefficient on wins and/or a positive coefficient on losses suggests a belief in the gambler's fallacy. A statistically insignificant coefficient is evidence for a belief in neither the hot hand myth nor the gambler's fallacy. The quadratic terms for wins and losses allow for non-linearities in these effects.

The cumulative amount won or lost in the trip allows us to control for the amount of money in the gambler's pocket. Inclusion of quadratic terms for cumulative wins/losses allows for non-linear effects. A particularly interesting aspect of this is that it allows us to find the threshold amount of wins/losses after which the sign of the total effect of wins/losses on play probability and bet amounts changes. This will allow us to test for loss aversion.

Comps are the primary marketing activity for the casino after the consumer has entered the casino. We include the dollar amount of comps in the trip to control for marketing activities.

The set of covariates thus consists of a combination of random outcomes, lagged behavioral variables and marketing activity and is given below.

$$X_{it} = \begin{pmatrix} 1 & Bet_{i,t-1} & Win_{i,t-1} & Win^2_{i,t-1} & Loss_{i,t-1} & Loss^2_{i,t-1} \\ & CumWin_{i,t-1} & CumWin^2_{i,t-1} & CumLoss_{i,t-1} & CumLoss^2_{i,t-1} & Comps_{it} \end{pmatrix} \quad (2)$$

To more clearly demonstrate how the win and loss variables are created, consider the following example. Suppose a person bets an amount of $100 on the first play in the trip and wins a total of $90 in that play. i.e. the net loss in the first play is $10. The first observation for the bet sub-model is the second play (since the inclusion of lagged covariates requires us to drop the first play). For this observation, the previous play had a loss of $10, so the $Loss_{i,t-1}$ and $CumLoss_{i,t-1}$ variables are set at 10 and the quadratic terms accordingly computed. All the Win variables are 0.

In the second bet, suppose the total amount bet by the customer was $80 and the amount won was $120, i.e. the second play generated a net 'win' of $40. In this observation, the variable $Loss_{i,t-1}$ is set to 0 and $Win_{i,t-1}$ is set to 40. $CumLoss_{i,t-1}$ is set to 0 and $CumWin_{i,t-1}$ set to 30.

Suppose the gambler has only one play in the trip. Then there is no observation for the bet sub-model. However, there is one observation for the play sub-model, which is described below.

The decision whether to play or not is assumed to be driven by an underlying latent variable that we term $Play^*_{it}$ given by the following equation

$$Play^*_{it} = Y_{it}\beta_i + \nu_{it} \tag{3}$$

subject to the following constraints

$$\begin{array}{ll} Play^*_{it} > 0 & \text{if } Play_{it} = 1 \\ Play^*_{it} < 0 & \text{if } Play_{it} = 0 \end{array} \tag{4}$$

Thus, (3) and (4) jointly describe the play sub-model.

Given this structure, for identification purposes, we need an instrument that affects the participation decision, but not the amount of the bet. The instrument we use is the length of time that the customer has played in the casino within the trip until that point of time. As time spent on a trip increases, the consumer has to make a decision whether to keep playing or not. The amount of time spent would be a proxy for satiation and/or fatigue and thus directly influence the decision to play. On the other hand, given that the consumer is going to play, it is unlikely that the amount of time spent playing so far affects the bet amount during the next play, controlling for the total amount of money won or lost until then. [11] Thus, the covariates for the

[11] We also confirmed this empirically i.e., we found that the correlation between the time spent in

play sub-model, Y_{it}, include all the covariates in the bet sub-model, X_{it}, in addition to the "excluded" variable, which is the time spent in the casino until t. Thus,

$$Y_{it} = \begin{pmatrix} 1 & Bet_{i,t-1} & Win_{i,t-1} & Win_{i,t-1}^2 & Loss_{i,t-1} & Loss_{i,t-1}^2 \\ & CumWin_{i,t-1} & CumWin_{i,t-1}^2 & CumLoss_{i,t-1} & CumLoss_{i,t-1}^2 & Comps_{it} \\ & TimeSpent_{it} & & & & \end{pmatrix} \tag{5}$$

As described earlier, we need to control for selection by modeling the correlation between the bet equation and play equation. Thus, we make the assumption that ε_{it} and ν_{it} have a joint bivariate normal distribution with a non-diagonal covariance matrix.

$$\begin{pmatrix} \varepsilon_{it} \\ \nu_{it} \end{pmatrix} \sim N\left(0, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix}\right) \tag{6}$$

Here, the correlation between the bet equation and the play equation is ρ. Note also that the variance of the marginal distribution of the play equation is fixed to 1 for identification purposes.

Thus, the bet sub-model is a log-linear regression and the play sub-model is a probit model. The play sub-model can also be thought of as a multi-period normal hazard model. The correlation ρ between these two sub-models explicitly controls for selection and we have an excluded variable, the length of time spent in the casino until t in the covariates for the play sub-model Y_{it} but not in the covariates for the bet sub-model X_{it}. Note also that all the variables that are in X_{it} can be in Y_{it}.

Combining 1, 3, 4 and 6, we can write the joint model as

$$\begin{array}{ll} \begin{pmatrix} \ln(Bet_{it}) \\ Play_{it}^* \end{pmatrix} \sim N\left(\begin{pmatrix} X_{it}\alpha_i \\ Y_{it}\beta_i \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix}\right) 1\left(Play_{it}^* > 0\right) & \text{if } Play_{it} = 1 \\ \qquad and & \\ Play_{it}^* \sim N\left(Y_{it}\beta_i, 1\right) 1\left(Play_{it}^* < 0\right) & \text{if } Play_{it} = 0 \end{array} \tag{7}$$

Thus, if the consumer decides to play, the model is a truncated bivariate normal

a trip and the amount bet within a play is very small, at 0.0829.

model, and a truncated univariate normal model otherwise. [12]

3.3 Hierarchical Model

Since the model is specified at the individual consumer level, we need to specify the hierarchy for the individual-level parameters. We assume that the individual level parameters in the bet and play sub-models for the play-level analysis are functions of observed demographic variables z_i and are distributed in the population as follows

$$\alpha_i \sim N\left(M_\alpha z_i, V_\alpha\right) \tag{8}$$

$$\beta_i \sim N\left(M_\beta z_i, V_\beta\right) \tag{9}$$

$$\gamma_i \sim N\left(M_\gamma z_i, V_\gamma\right) \tag{10}$$

The model is completed by specifying the prior distributions for all the population-level parameters (σ^2,ρ,M_α, V_α, M_β, V_β).

$$\sigma^2 \sim \frac{\nu_\sigma s_\sigma^2}{\chi^2_{\nu_\sigma}} \tag{11}$$

$$\rho' = \frac{\rho+1}{2} \sim \text{Beta}\left(\lambda_1, \lambda_2\right) \tag{12}$$

$$\sigma_\xi^2 \sim \frac{\nu_\xi s_\xi^2}{\chi^2_{\nu_\xi}} \tag{13}$$

$$\text{vec}\left(M_\alpha\right)|V_\alpha \sim N\left(\bar{\alpha}, V_\alpha \otimes A^{-1}\right) \tag{14}$$

[12]It is important to note here that this specification gives us valid results under the assumption that there is no serial correlation in the unobserved term. If such serial correlation exists, it has potential to bias the results particularly in the case of the bet sub-model, where we have a lagged dependent variable. We rule out autocorrelation in the residuals of the bet sub-model (at the mean levels of the parameters) using Durbin-Watson and Breusch-Godfrey tests, thus mitigating concerns about such a bias.

$$V_\alpha \sim \text{Inverse Wishart}\left(\mu_\alpha, S_\alpha\right) \tag{15}$$

$$\text{vec}\left(M_\beta\right)|V_\beta \sim N\left(\bar{\beta}, V_\beta \otimes B^{-1}\right) \tag{16}$$

$$V_\beta \sim \text{Inverse Wishart}\left(\mu_\beta, S_\beta\right) \tag{17}$$

$$\text{vec}\left(M_\gamma\right)|V_\gamma \sim N\left(\bar{\gamma}, V_\gamma \otimes G^{-1}\right) \tag{18}$$

$$V_\gamma \sim \text{Inverse Wishart}\left(\mu_\gamma, S_\gamma\right) \tag{19}$$

Here, A, B and G are the matrices of stacked α_i, β_i and γ_i vectors respectively. The demographic variables, z_i, we include are gender, age, racial/ethnic identity and an indicator for whether the consumer reports that he/she prefers to play table games or slots.

3.4 Estimation

As mentioned in the data section, we estimate our within trip play-level model on a random sample of 2000 customers. For the sake of estimation, we lose one observation per trip, since we include the previous bet amount in the specification to control for addiction. This leaves us with 40510 observations, although we have 56142 observations for the play sub-model (since there is one last observation of the consumer not choosing to play). Within a trip, the vector of $Play_{it}$ looks like the following

$$Play_{it} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & ... & 0 \end{pmatrix}' \tag{20}$$

The estimation of this model requires us to draw from the joint posterior distribution of all parameters conditional on the data. Since the joint distribution is not from a known family of distributions and cannot be drawn from directly, we use the Gibbs Sampler (Gelfand and Smith 1990), making sequential draws from the full conditional distributions of sub-vectors of the parameter vector. At convergence, these sequential draws from the full conditional distributions give us draws from the joint distribution of the parameter vector. Further, we use data augmentation (Tanner and Wong 1987) and draw from the full-conditional distribution of $Play^*_{it}$, taking into account the fact that it is a truncated normal distribution.

The derivations of the full conditional distributions of the various parameters are given in the Appendix. Note that the full conditional distributions of σ^2 and ρ cannot be directly drawn from. Hence, we use the Random Walk Metropolis-Hastings algorithm (Chib and Greenberg 1995) to make draws from the full conditional distributions of these parameters. The remaining sub-vectors of the parameter vector are drawn directly.

The chain was run for a total of 30,000 draws, with the first 25,000 draws being discarded and the inference done on the remaining 5,000 draws. Convergence was assessed by inspecting the chains for each of the parameters.

4 Results

We now discuss our results. We first describe the parameter estimates for the individual level parameters of interest. We then look at the economic significance of these parameters, examine the correlation patterns between these parameters and describe how observed demographics explain individual-level variation in behavior.

4.1 Parameter Estimates

First, we look at the individual-level parameters, which are reported in Table 4.[13] Recall that it is crucial for us to be able to estimate the individual parameters with a reasonable degree of certainty so that we can test for the various effects that we have outlined. Our approach allows us to compute the posterior distribution of each parameter for each individual. A natural question to ask is if the posterior distribution

[13] We report the individual-level parameter estimates for 1769 out of 2000 consumers in the estimation sample. For the remaining 231 consumers, the parameters of the bet equation are not defined since these are households for which there were no trips with more than one play. We have play equation parameters for these households, but to ensure consistency, we report play equation parameters also for the same 1769 consumers.

of these parameters are tightly estimated. We use the ratio of the posterior mean to the posterior standard deviation as our informal metric to determine if each individual parameter is "significant." Our cutoff to determine this significance is 1.65 (akin to the t-statistic at p=0.10). We report the distribution of the means of the individual-level parameters, i.e. the mean, standard deviation, median and other percentiles for the individual-level mean for each parameter. The coefficient for $LastBet$ is negative on average for both the bet sub-model and the play sub-model. We find that the coefficient for $LastBet$ is negative for most of the households in both the bet and play sub-model, although it is significantly negative at the 10% level or higher for 525 and 39 consumers respectively for the bet sub-model and play sub-model. It is positive and significant for the bet sub-model for 12 consumers and for 148 consumers for the play sub-model. Thus, for a majority of consumers in the sample we do not find evidence for addiction to gambling. 12 consumers display evidence for addiction as tested by their bet amounts, while 148 consumers display some evidence for addiction in terms of a positive effect of the previous bet amount on the probability of playing again in the trip. These constitute 0.1% and 8.4% of the sample respectively. [14] All 12 individuals who have a significantly positive coefficient for the bet sub-model also have significantly positive coefficients for the play sub-model. Again, note that these results should be interpreted in the context of our definition of addiction, which predicts a positive coefficient for $LastBet$.

It is interesting to note that this proportion of consumers who display some evidence of addictive behavior in our analysis is comparable to that reported in the medical literature based on self-reported surveys of gamblers.[15]

The coefficients for $LastWin$ and $LastWin^2$ are negative and positive respectively for both the sub-models, but are largely not significant at the individual level for the bet sub-model. While there is no significant effect of $LastWin$ on bet amounts for any of the consumers in the sample, there is a significantly negative effect on the probability of subsequent play for over 95% of the consumers. Thus, if a consumer

[14] As a robustness check, we added another randomly chosen 2000 customers to our estimation sample and ran the model. We find that over 80% of the customers who were identified as addicted in the estimation sample were also classified as addicted based on the results obtained using the augmented sample.

[15] See, for instance, Potenza, Kosten, and Rounsaville (2001) who use survey data and report prevalence rates for addiction of about 5% amongst casino gamblers. Our results provide some support for these estimates via the use of behavioral data.

has a win in the previous occasion, it increases her probability of not playing again in the trip. On average, the coefficients for $LastLoss$ and $LastLoss^2$ are negative and positive respectively for the play sub-model, but are insignificant for all consumers. For the bet sub-model, these coefficients are positive and negative respectively, though the latter is not significant for any consumer. The coefficient for $LastLoss$ is positive and significant for 53 consumers in the bet equation. All of these results are consistent with these consumers maintaining the gambler's fallacy, since they suggest a negative correlation between the previous outcome and the current propensity to continue playing and the amount bet. This is an important set of results since it is, to our knowledge, the first study to document the gambler's fallacy amongst casino gamblers, using behavioral data and confirms the results of previous studies using laboratory experiments or direct observations of small samples of consumers (Croson and Sundali 2005). It is also opposite to the results of Guryan and Kearney (2008), who document the presence of a hot hand myth using aggregate data for state lotteries.

The variable $CumWin$ has an insignificant effect on bet amounts for all consumers, but has an interesting effect on the probability that consumers will continue to play. On average, $CumWin$ has a positive coefficient and $CumWin^2$ has a negative coefficient. An overwhelming majority of consumers - 1747 and 1657 respectively - have a positive effect of $CumWin$ and a negative effect of $CumWin^2$ on the subsequent play. Thus, consumers continue to play for small wins, but seem to cash out after their cumulative wins have reached a high level. We will discuss heterogeneity in this cashing out behavior later in this section.

Another interesting effect is that of $CumLoss$ on subsequent play. The coefficient for $CumLoss$ is positive and that for $CumLoss^2$ is negative on average, and significant for a vast majority of consumers (1721 and 1633 consumers respectively). Thus, consumers seem to continue to play on as long as the losses are small, but stop playing when losses exceed a certain amount. This seems to suggest that consumers set a certain limit up to which they tolerate losses, but after that, they cut their losses by terminating play (more details on these limits are given below).

Both these results are significant, since they seem to indicate, albeit informally, that consumers treat casino gambling like a form of entertainment and behave somewhat rationally on average, cashing out when the "going is good" and cutting their losses when they are losing a lot.

The effect of *Comps* is positive and significant for both the bet and play sub-models for a majority of consumers. Thus, it appears that marketing activity is positively related to both the bet amount and the probability of participation. Finally, in the play sub-model, the excluded variable, which is time of play, has a negative coefficient. As discussed before, this may be due to satiation, play fatigue or time constraints in general.

Table 4 about here

4.2 Elasticities

The parameter estimates discussed above are useful to understand the directional nature of the various effects we have discussed. In order to understand their magnitudes, we compute elasticities for various factors that affect the play and participation decisions and report them in Table 5. These elasticities are computed for the mean consumer and are computed for the covariates for each observation in our data and measure the magnitude of the effect of the covariates, *keeping everything else constant*. We then take the average across these observations. We report three sets of elasticities for each covariate. The first is the elasticity of the bet amount, conditional on participation. The second is the elasticity of the probability of participation. The third is the total elasticity, which combines the participation and bet amount and is hence the unconditional elasticity of the bet amount.

The unconditional bet amount elasticity for $Comps$ is about 0.14. The bulk of this effect is from the impact of comps on the amount bet (0.11), with a smaller impact on the probability of continuing play.

The elasticity of $LastBet$ is negative and fairly large for the average consumer at -0.28. The major component of this elasticity is from the bet equation. The elasticity of $LastWin$ is relatively small, with the bet amount elasticity being insignificant because of insignificant parameter estimates. $LastLoss$ has a positive overall elasticity, with a positive effect on bet amounts and negative effect on the probability of play.

The elasticity of $CumWin$ is relatively small, while that of $CumLoss$ is relatively large. $CumWin$ has a positive effect on bet amounts as well as play probabilities, while *CumLoss* has a negative bet amount elasticity and a positive play elasticity, leading to a negative overall elasticity. The overall elasticity, at about 0.25, is at almost the level of the *LastBet* elasticity, and is almost twice that for *Comps*.

Table 5 about here

4.3 Heterogeneity

We next examine heterogeneity in the parameters of interest and try to relate this heterogeneity with the observed characteristics of consumers. In our Hierarchical Bayesian approach, we do this by specifying a level of the hierarchy that relates individual-level parameters to observed demographic and other characteristics. The observed characteristics in our dataset are gender, age, race and self-reported preference for slots versus table games. The parameters for this level of the hierarchy are reported in table 6. We discuss only the significant parameter estimates (indicated with stars in the table), with significance defined as before (i.e., significance at 90% level being defined as the ratio of posterior parameter mean and the posterior standard deviation being greater than 1.65 in absolute value).

First looking at the addiction parameter - the coefficient for *LastBet* - for the bet equation, younger consumers and those who have a stated preference for table games are more likely to be addicted, while Hispanics are less likely to be addicted than other ethnic groups. In terms of the effect on play, males and Hispanics are slightly less likely to be addicted. Comparing the 148 consumers who are classified as addicted, based on the sign and statistical significance of their *LastBet* coefficients with the full set of consumers, we find that there is a greater proportion of women (60.8% versus 54.1% for the overall sample) of Western European origins (68.2% versus 64.6%). In terms of activity, this is a group that is somewhat more active in gambling with slightly more trips (10.7 trips versus 8.4), more plays on a trip (6.4 versus 4.2) and heavier amounts bet on a trip (mean amount bet per trip at $ 661 versus $ 469). On average, the number of days between trips was lower (68 versus 85) and the amount of time spent per play was higher (87 minutes versus 72 minutes).

The coefficients for *LastWin* and *LastLoss* test for the gambler's fallacy and the hot hand myth. Hispanic and male consumers seem to more strongly display evidence for the gambler's fallacy in terms of their betting behavior, although not in terms of their participation decision. Older consumers are also more likely to show evidence of the gambler's fallacy in terms of the effect of $LastLoss$ on bet amounts.

In general, we find that the magnitude of the demographic effects is relatively small, i.e., the variation in the response parameters characterizing gambling behavior is explained largely by unobservables rather than the observables (demographics).

Table 6 about here

An interesting result of our empirical analysis is that for most consumers, the probability of continuing play is concave in winnings, i.e. the probability increases in winnings up to a point and then starts reducing. We can thus quantify the maximum levels of winnings and losses for the consumer after which the marginal effect of previous wins and losses on the participation decision is negative. The median consumer has a maximum wins level of $ 941.60 and a maximum loss level of $ 382.60. The 5th percentile levels for wins and losses are $ 520.50 and $ 229.60 respectively, while the 95th percentile levels are $ 1671.50 and $ 1322.50 respectively.

It is also interesting to look at the ratio of wins and losses. Prospect theory predicts that due to loss aversion, consumers are likely to cash out and quit playing after a smaller level of losses than wins. This is corroborated by our estimates. On average, consumers are willing to take fewer losses than wins before that starts negatively affecting their participation decision. The ratio of wins to losses after which they respectively start to have a negative effect is 2.47 at the median (of the ratio distribution). Previous experimental research has found this ratio to typically be between 2 and 4 (Heath, Larrick, and Wu 1999). This is an important finding in that ours is one of the first studies to document this prediction of prospect theory on individual-level behavioral data, especially in the gambling domain. Note also that the ratio distribution also exhibits wide variation - the 5th and 95th percentile values for this ratio are 1.12 and 6.49 respectively.

4.4 Parameter correlations

The coefficients for *LastBet* and *Comps* are positively correlated in the play sub-model, with a correlation of 0.63.[16] This provides some directional evidence for the fact that consumers who are more prone to addictive behaviors are also more strongly affected by marketing activities in this category. To illustrate this, we present the scatter plots of the individual-level mean estimates for these two coefficients for the play sub-model in Figure 1.

We also find a high level of correlation between the coefficient for *LastBet* and the intercept in both the bet and play sub-models (0.84 and 0.83 respectively). This suggests that consumers who have a high intrinsic bet amount and a high intrinsic probability of participation are also more likely to be prone to addictive behaviors. This can be also seen in the scatter plots in figures 2 and 3.

Figures 1 - 3 about here

5 Marketing and Addiction

One of the important issues that we try to investigate in this research is the role played by marketing in this industry and especially its interaction with addicted gamblers. We have already seen that the elasticity of comps is about 0.11 in the case of the amount bet, and about 0.03 in the case of play probabilities. We have also seen that individuals who are likely to be addicted are more responsive to comps as well (the individual level coefficients show a positive correlation of 0.63). Thus, comps affect addicted individuals more than non-addicted individuals. The comps elasticity for the addicted individuals is 0.13 and 0.035 for the amount bet and play probabilities respectively.

These short run elasticities, while interesting, do not paint a complete picture of the effect of comps on the addicted individuals. In addition to the direct effect of comps on the level of play, there is an indirect effect due to addiction. For all gamblers,

[16] These two coefficients exhibit a negative correlation in the bet sub-model. However, we focus on the play sub-model as many more of the *LastBet* parameters are significant in that sub-model relative to the bet sub-model.

our results show that comps increase the amount bet within a play. However, for addicted gamblers, this increased level of betting in turn results in an increase in the amount bet in the next play and increases the probability that the individual will continue to play. As a result of this, the long-run effect of comps has an additional multiplier due to the effect of addiction for these individuals.

We explore this addiction multiplier of the effect of comps via simulation. First, we fix the variables such as $LastBet$, $LastWin$, $LastLoss$, $CumWin$, $CumLoss$ and $TimeSpent$ at the average values after the first bet across all consumers i.e. we initialize these variables to the averages after the first bet. Next, we simulate the bet amount and play decision (using the estimated play probabilities). We also simulate the wins/losses in that play by using the mean win/loss proportions across the data (i.e., using the empirical distribution of the casino's win/loss ratios). Finally, we assume that the time spent in the play is equal to the average time spent across plays in the data. If the simulated decision is to terminate the trip, we stop the simulation run, recording the total amount bet across all the plays in the simulated trip and the total number of plays in that trip. We repeat this simulation 1000 times. Next, we increase the comps by a small amount (1%) and repeat the whole exercise. We report the elasticities of comps using these simulation runs by computing the percentage change in the total amount bet per trip and the number of plays during the trip.

We repeat these computations three times and compute elasticities thrice, for three different values of the parameters. For the first set of elasticities, we fix the parameters of the bet and play equations to the average levels of the parameters across all individuals. In other words, these elasticities are for a representative consumer in our sample. For the second set of simulations, we fix the parameters to the mean levels of the parameters for the individuals who are classified as addicted, based on the coefficient for $LastBet$ in the play equation. Finally, we fix the parameters to the mean levels for all other individuals, i.e., everybody who is not classified as addicted. These estimates are reported in Table 7.

At the estimated parameter levels, the elasticities are somewhat lower than the short-run elasticities. This reflects the fact that the estimates for the coefficient of $LastBet$ are negative for most consumers. Thus, a higher bet due to comps results in fewer future plays and a lower amount bet in the future, making the long run elasticities lower than the short-run ones. The next two sets of elasticities are more interesting, since they allow us to compare the long-run effect of comps for addicted versus non-addicted individuals. We find that the addicted individuals have a long-run comp elasticity on the total amount bet of about 0.18 versus 0.08 for those not classified as addicted. In terms of number of plays, the elasticities are 0.16 and 0.08 respectively. Thus, the elasticities for the addicted individuals are significantly higher (at least twice) than those for the non-addicted individuals.

In summary, our results show that the long-run effect of comps is lower than the short-run effect for the average gambler, as most gamblers are not addicted. Thus, comps may increase the amount bet, but this results in a lower amount bet in the next play and a higher probability of terminating the trip. In other words, firms need to think about the tradeoff between the (larger) short-run and (smaller) long-run benefits of comps for the typical gambler when they set comp levels. Second, and perhaps more important, addiction causes comps to have a multiplier effect for addicted gamblers i.e., for addicted gamblers, both the short-run and long-run effects of comps are in the same direction. Thus, if the casino uses long-term gain as a criterion in terms of targeting gamblers for marketing resource allocation, they may end up (inadvertently) focusing on addicted gamblers and therefore drawing the attention of regulators. These results are also likely to be helpful to policymakers as they quantify the effects of marketing on different gambler populations.

Table 7 about here

6 Robustness Check: Across Trip Analysis

As alluded to in section 3.1, we check the robustness of our findings using an alternate unit of analysis. Specifically, in contrast to our main (within-trip) analysis, we use a trip itself as the unit of analysis. Conceptually speaking, a likely manifestation of addictive behavior is acceleration in trips. This is because one of the processes by which addiction operates is through the phenomenon of "tolerance" (Peele 1985, Becker and Murphy 1988). Tolerance is the phenomenon by which a given level of consumption provides lower utility to the consumer as she becomes more addicted. A consequence of this phenomenon is that the consumer is less satiated with a given level of consumption and this lowered satiation causes the consumer to want to consume the product (in this case gambling) sooner after the last consumption occasion. In our data, this will manifest itself as a negative effect of previous consumption level on the inter-trip duration.

Thus, the across-trip analysis tests for whether the inter-trip duration is affected by the amount of money gambled in the previous trip. Once again, controlling for heterogeneity is important in this context to allow for correlation between the inter-trip duration and amount gambled. For instance, it is possible that consumers with shorter inter-trip durations also happen to, on average, bet higher amounts in each trip, leading to a spurious effect of the amount gambled in the previous trip on inter-trip duration unless such heterogeneity is controlled for. We conduct this inter-trip analysis using a log-linear regression model, where the dependent variable is the log of the duration between one trip and the next and the econometric error ξ_{it} is assumed to be normally distributed with mean 0 and variance σ^2_ξ: [17]

$$ln\left(duration_{it}\right) = W_{it}\gamma_i + \xi_{it} \tag{21}$$

The key covariate in the specification above is the total amount bet in the previous trip. Another important factor that could affect the inter-trip duration is the budget constraint of the consumer, which could lead to a positive effect of past consumption on the inter-trip duration. If consumers bet more in a certain trip, they would also lose more money in expectation. If they face a budget constraint, this would also mean that they would not be able to return to the casino as quickly as they would have if they had lost less money. While we do not have a direct measure of the

[17] An alternative approach would be to specify a hazard model, where the probability of visiting the casino given that the consumer has not visited it in the weeks since the previous trip is modeled as a hazard function, with the hazard being modified by the consumption level (i.e. total amount bet) in the previous trip. We chose the linear model because of its simplicity and its ability to directly test for the addiction effect. Furthermore, the hazard model would need us to specify the unit of time for an observation of whether the consumer visited the casino or not. In the case of grocery products, it is reasonable to assume that consumers plan their shopping on a weekly basis, and hence the natural observation is at a weekly level. However, in the case of casinos, it is unclear what this unit of observation should be. Trips extend across multiple days in many cases, and hence a day cannot be a reasonable unit of analysis. In addition, a single trip may spill over the weekly boundary.

budget constraint faced by the consumer, we include the total amount won or lost in the previous trip as covariates in the regression to control for this effect. Thus, we would measure the effect of past consumption in a particular trip on the duration until the next trip, controlling for the amount won or lost in the trip.

A control for heterogeneity is also important, else we would spuriously pick up heterogeneity as an addiction effect. If consumers are heterogeneous in terms of their inter-trip duration, with some consumers visiting the casino more frequently than others, and if this is correlated with their level of betting in each trip, the model would pick up this correlation as a positive effect of past consumption on inter-trip duration if we were to not control for this heterogeneity. Hence, we include a rich specification of heterogeneity at the individual level in our specification.

We also test for the effects of the casino's marketing program on inter-trip duration by including the total amount of comps in the previous trip as a covariate. Finally, we include total time spent gambling in the t^{th} trip in order to control for satiation effects. The specification for the inter-trip analysis is given below

Here i indexes the individual consumer and t indexes the trip. The dependent variable $duration_{it}$ is the duration between the t^{th} trip and the $(t+1)^{th}$ trip. The covariate vector W_{it} is given as follows, with all the variables being the totals of the respective variables for all plays within the t^{th} trip.

$$W_{it} = \begin{pmatrix} 1 & Bet_{it} & Win_{it} & Win_{it}^2 & Loss_{it} & Loss_{it}^2 & Comps_{it} & Time_{it} \end{pmatrix} \tag{22}$$

6.1 Sample

We conduct this analysis for two samples from our database. The first sample is the same (random) sample that we use for the within-trip analysis. However, as described earlier, an across-trip analysis is likely to be affected by the nature of our data. Therefore, we also carry out this analysis for a sample of consumers that visits the casino much more frequently. Sampling based on this criterion helps us to lower the probability that these consumers visit other casinos in the period between two visits to the focal casino. It also has the additional benefit of providing us adequate data points per individual consumer, thus allowing the model the best chance of recovering the individual level estimates for the covariates that drive across-trip behavior.

This second sample, consisting of 2000 more "frequent" gamblers, is drawn from the universe of customers who made at least six trips to the casino is a year. Thus, these were consumers who made at least one trip to the casino at least every two months on average. Table 2 compares the behavioral information about these consumers with those for the random sample, and Table 3 compares the demographic information for the random consumer and the set of more frequent visitors to the casino used for the inter-trip analysis. As expected, in terms of activity, relative to the random sample, this sample consists of more frequent visitors who spend more time and money on gambling. In terms of demographic comparisons of the random sample and frequent gamblers' sample, the most significant differences are in the proportions of Hispanics (about 18% for the frequent gamblers vs. 10% for the random sample) and lower preference for slot machines relative to table games for the frequent gamblers (66.2% vs. 85.9% for the random sample).

6.2 Results

The mean parameter estimates for the trip-level analysis for both samples are given in Table 8 and Table 9. The main parameter of interest to us is the coefficient on the variable Bet_{it} - the total amount of money that the consumer gambled in the previous visit to the casino. A negative sign on this parameter is indicative of addiction at the inter-trip level. As described earlier, this results from tolerance to gambling with increased addiction levels, resulting in the consumer being less satiated with a given level of gambling and consequently a shorter duration until the next visit to the casino.

For the frequent gamblers' sample (Table 9) , we find that a total of 362 consumers amongst the sample of 2000 consumers have a negative and significant coefficient (at the 90% level) for the Bet_{it} variable, indicating addictive effects for these consumers, accounting for 18.1% of the sample. For the random sample, we find that the the coefficient for the Bet_{it} variable is not negative and significant for any consumer. In other words, the across-trip results suggest that the presence of addiction for a substantial proportion of frequent gamblers but not for the random sample. Given our discussion, the latter results is not unexpected.

Tables 8 and 9 about here

Finally, in order to check that our definition of addiction is consistent across behaviors (i.e., within and across trips), we examined the correlation between the coefficient of the previous bet amount for the across-trip analysis (negative for addicted consumers) and the coefficient of the previous bet among for the play equation (positive for addicted consumers) for both samples.[18] We find that the individual level addiction parameters for the two models have a negative correlation of 0.71 for the frequent gamblers' sample and 0.62 for the random sample of gamblers (although the across-trip parameters are not significant for any consumer for the random sample). Thus, consumers who are prone to addiction within trips in terms of their tendency to continue playing are also likely to accelerate their next trip to the casino due to addiction for both samples. Figures 4 documents the correlations for the frequent gamblers' sample via a scatter plot.

Figure 4 about here

7 Conclusion

Our paper adds to the small but growing body of research that investigates gambling behavior. As noted before, the gambling industry is a large and pervasive part of American society today. This has led to a lot of debate on the benefits and costs of gambling. Our access to a rich and new dataset on individual consumer behavior vis-a-vis casino visitation and activity allows us to take a data based approach to investigating many of the commonly made arguments about these benefits and costs. Specifically, we investigate three charges frequently leveled against the gambling industry – it leads to addictive behavior (with potentially harmful individual and societal effects), it leverages irrational consumer beliefs and it uses marketing incentives to influence gamblers.

[18]We also conducted a within-trip analysis for the sample of frequent gamblers. We found that, for this sample, 19.1% of consumers are classified as displaying addictive behaviors based on the "play" equation and 6.3% based on the "bet" equation. Detailed results for this model are available from the authors on request.

We use the commonly accepted definition of addiction from the economics literature to test for its presence i.e., that current consumption is affected by past consumption. We fit a model of the play decision and bet amount (given play) to data from a consumer panel of casino visitors over a two year period. Our data are at a highly disaggregate level – we look at play decisions within a given trip for individual consumers. Our modeling approach allows us to exploit the rich variation in the data both across and within individuals.

Our results show that, controlling for other reasons that could induce play, only about 8% of all consumers show evidence for addiction (as defined by us). While this proportion may look small (in absolute terms), it is consistent with research in other academic fields that has focused on casino gamblers. This result also suggests, that for a majority of casino gamblers, the failure to find patterns of addiction may be interpreted as support for the view that the role of casinos for these gamblers is to provide entertainment. We also check the robustness of our findings using a different level of analysis (the trip itself) and find that the extent of addiction goes up as we move from a random sample of gamblers to a sample that visits the casino frequently.

An important aspect of our analysis is that we conduct an empirical test for two kinds of irrational beliefs - the hot hand myth and the gambler's fallacy. We find evidence for the gambler's fallacy in both directions. For a subset of consumers we find that if they win a bet, they are less likely, on average, to continue betting. The opposite - if consumers have lost the previous bet, the likelihood that they would continue betting increases - also holds true, albeit for a smaller number of consumers. We believe this is the first study to conduct such an analysis on individual-level behavioral data. Previous research has documented such effects using either laboratory data, or aggregate across-consumer data, which is prone to alternative accounts. We also find evidence consistent with loss aversion in terms of the decision to continue playing as a function of previous winnings and losses.

Our paper also helps to quantify the effects of marketing on gamblers. We find that marketing activity has a short-run positive effect on the decision to play and the amount to bet. We also find that addicted gamblers are more responsive to marketing than non-addicted gamblers. This leads to a long-run effect of marketing activities that is twice as large for addicted gamblers than others. Finally, while for a typical (non-addicted) gambler, the (larger) short-run effect of marketing is counterbalanced by the (smaller) long-run effect, this is not true for addicted gamblers. These results have implications for the firm and policy makers.

In terms of continuing research, we see a lot of scope for refining our approach. First, our results are based on our definition of addiction as well as our choice of the level of disaggregation. Second, we have assumed that consumers are myopic. As discussed earlier, estimation of an individual-level model with forward-looking behavior is challenging in this context and represents an avenue for further research. Finally, modeling choices could be even better informed through the use of institutional details. For example, as described earlier comps and offers are targeted strategically. While our rich specification of heterogeneity controls for this, another approach would be to explicitly model the non-strategic nature of comps.

References

AYTON, P., AND I. FISCHER (2004): "The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness?," *Memory and Cognition*, 32(8), 1369–1378.

BECKER, G. S., M. GROSSMAN, AND K. M. MURPHY (1994): "An Empirical Analysis of Cigarette Addiction," *American Economic Review*, 84(3), 396–418.

BECKER, G. S., AND K. M. MURPHY (1988): "A Theory of Rational Addiction," *Journal of Political Economy*, 96(4), 675–700.

CHALOUPKA, F. (1991): "Rational Addictive Behavior and Cigarette Smoking," *Journal of Political Economy*, 99(4), 722–742.

CHIB, S., AND E. GREENBERG (1995): "Understanding the Metropolis Hastings Algorithm," *American Statistician*, 49, 327–335.

CLOTFELTER, C. T., AND P. J. COOK (1993): "The Gambler's Fallacy in Lottery Play," *Management Science*, 39(12), 1521–1525.

CROSON, R., AND J. SUNDALI (2005): "The Gambler's Fallacy and the Hot Hand: Empirical Data from Casinos," *Journal of Risk and Uncertainty*, 30(3), 195–209.

GELFAND, A. E., AND A. F. M. SMITH (1990): "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85(410), 398–409.

GILOVICH, T., R. VALLONE, AND A. TVERSKY (1985): "The Hot Hand in Basketball: On the Misperception of Random Sequences," *Cognitive Psychology*, 17, 295–314.

GRUBER, J., AND B. KOSZEGI (2001): "Is Addiction Rational? Theory and Evidence," *Quarterly Journal of Economics*, 116(4), 1261–1303.

GURYAN, J., AND M. S. KEARNEY (2005): "Lucky Stores, Gambling and Addiction: Empirical Evidence from State Lottery Sales," NBER Working Paper 11287.

——— (2008): "Gambling at Lucky Stores: Empirical Evidence from State Lottery Sales," *American Economic Review*, 98(1), 458–473.

HEATH, C., R. P. LARRICK, AND G. WU (1999): ""Goals as Reference Points"," *Cognitive Psychology*, 38, 79–109.

HECKMAN, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–161.

——— (1981): "Statistical Models for Discrete Panel Data," in *Structural Analysis of Discrete Data and Econometric Applications*, ed. by C. F. Manski, and D. L. McFadden. MIT Press.

Jones, J. (2004): "Profile of the American Gambler," available at http://www.harrahs.com/about_us/survey/2004_Survey.pdf, last accessed on June 15, 2006.

Kahneman, D., and A. Tversky (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 263–291.

Kearney, M. S. (2005): "The Economic Winners and Losers of Legalized Gambling," *National Tax Journal*, 58, 281–302.

Metzger, M. (1984): "Biases in Betting: An Application of Laboratory Findings," *Psychological Reports*, 56, 883–888.

Mullainathan, S. (2002): "Thinking Through Categories," Working Paper, Harvard University.

Olekalns, N., and P. Bardsley (1996): "Rational Addiction to Caffeine: An Analysis of Coffee Consumption," *Journal of Political Economy*, 104(5), 1100–1104.

Peele, S. (1985): *The Meaning of Addiction: Compulsive Experience and Its Interpretation*. Lexington, Mass.

Pollak, R. A. (1970): "Habit Formation and Dynamic Demand Functions," *Journal of Political Economy*, 78(4), 745–763.

Potenza, M. N., T. R. Kosten, and B. J. Rounsaville (2001): "Pathological Gambling," *Journal of the American Medical Association*, 286(2), 141–144.

Roy, R., P. K. Chintagunta, and S. Haldar (1996): "A Framework for Investigating Habits: "The Hand of the Past," and Heterogeneity in Dynamic Demand Choice," *Marketing Science*, 15(3), 280–299.

Stigler, G. J., and G. S. Becker (1977): "De Gustibus Non Est Disputandum," *American Economic Review*, 67(2), 76–90.

Tanner, T., and W. Wong (1987): "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528–549.

Terrell, D. (1994): "A Test of the Gambler's Fallacy: Evidence from Parimutuel Games," *Journal of Risk and Uncertainty*, 8, 309–317.

Tversky, A., and D. Kahneman (1971): "Belief in the law of small numbers," *Psychological Bulletin*, 76, 105–110.

A Appendix A: Likelihood Function

For ease of exposition, we use the following abbreviations in this appendix

$$b_{it} = \ln\left(Bet_{it}\right) \tag{A-1}$$

$$P_{it} = Play_{it} \tag{A-2}$$

$$P^{*}_{it} = Play^{*}_{it} \tag{A-3}$$

Also, define the following

$$d_{it} = \begin{pmatrix} b_{it} \\ P^{*}_{it} \end{pmatrix} \tag{A-4}$$

$$g_{it} = \begin{pmatrix} X_{it}\alpha_i \\ Y_{it}\beta_i \end{pmatrix} \tag{A-5}$$

$$\Sigma = \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix} \tag{A-6}$$

$$m_\alpha = \text{vec}\left(M_\alpha\right) \tag{A-7}$$

$$m_\beta = \text{vec}\left(M_\beta\right) \tag{A-8}$$

The likelihood function can be written as follows

$$\mathcal{L} \propto \prod_{i=1}^{N} \left[\begin{array}{c} \prod_{t=1}^{T_i} \left[\begin{array}{c} \left[|\Sigma|^{-0.5} \exp\left(-\frac{1}{2}\left(d_{it} - g_{it}\right)' \Sigma^{-1} \left(d_{it} - g_{it}\right)\right)\right]^{P_{it}} \\ \left[\exp\left(-\frac{1}{2}\left(P^{*}_{it} - Y_{it}\beta_i\right)^2\right)\right]^{1-P_{it}} \\ \left[1\left(P^{*}_{it} > 0\right)\right]^{P_{it}} \left[1\left(P^{*}_{it} < 0\right)\right]^{1-P_{it}} \end{array} \right] \\ |V_\alpha|^{-0.5} \exp\left(-\frac{1}{2}\left(\alpha_i - M_\alpha z_i\right)' V_\alpha^{-1} \left(\alpha_i - M_\alpha z_i\right)\right) \\ |V_\beta|^{-0.5} \exp\left(-\frac{1}{2}\left(\beta_i - M_\beta z_i\right)' V_\beta^{-1} \left(\beta_i - M_\beta z_i\right)\right) \end{array} \right] \tag{A-9}$$

The joint posterior distribution of all parameters is proportional to the product of the likelihood and the prior densities and is given by

$$f\left(\text{parameters}|\text{data}\right) \propto \begin{array}{c} \prod_{i=1}^{N} \left[\begin{array}{c} \prod_{t=1}^{T_i} \left[\begin{array}{c} \left[|\Sigma|^{-0.5} \exp\left(-\frac{1}{2}\left(d_{it}-g_{it}\right)' \Sigma^{-1}\left(d_{it}-g_{it}\right)\right)\right]^{P_{it}} \\ \left[\exp\left(-\frac{1}{2}\left(P_{it}^{*}-Y_{it}\beta_i\right)^2\right)\right]^{1-P_{it}} \\ \left[1\left(P_{it}^{*}>0\right)\right]^{P_{it}}\left[1\left(P_{it}^{*}<0\right)\right]^{1-P_{it}} \end{array} \right] \\ |V_\alpha|^{-0.5}\exp\left(-\frac{1}{2}\left(\alpha_i-M_\alpha z_i\right)' V_\alpha^{-1}\left(\alpha_i-M_\alpha z_i\right)\right) \\ |V_\beta|^{-0.5}\exp\left(-\frac{1}{2}\left(\beta_i-M_\beta z_i\right)' V_\beta^{-1}\left(\beta_i-M_\beta z_i\right)\right) \end{array} \right] \\ |V_\alpha\otimes A^{-1}|^{-0.5}\exp\left(-\frac{1}{2}\left(m_\alpha-\bar{\alpha}\right)'\left(V_\alpha\otimes A^{-1}\right)^{-1}\left(m_\alpha-\bar{\alpha}\right)\right) \\ |V_\beta\otimes B^{-1}|^{-0.5}\exp\left(-\frac{1}{2}\left(m_\beta-\bar{\beta}\right)'\left(V_\beta\otimes B^{-1}\right)^{-1}\left(m_\beta-\bar{\beta}\right)\right) \\ |V_\alpha|^{-\frac{\mu_\alpha+k_1+1}{2}}\operatorname{etr}\left(-\frac{1}{2}S_\alpha V_\alpha^{-1}\right)|V_\beta|^{-\frac{\mu_\beta+k_1+1}{2}}\operatorname{etr}\left(-\frac{1}{2}S_\beta V_\beta^{-1}\right) \\ \left(\sigma^2\right)^{-\left(\frac{\nu_0}{2}+1\right)}\exp\left(-\frac{\nu_0 s_0^2}{2\sigma^2}\right)\left(\frac{1+\rho}{2}\right)^{\lambda_1-1}\left(\frac{1-\rho}{2}\right)^{\lambda_2-1} \end{array} \tag{A-10}$$

B Appendix B: Full Conditional Distributions

The full conditional distributions for each parameter vector is obtained by taking out all the terms from the joint posterior distribution in A-10, since the other terms affect only the proportionality constant. We inspect these terms to see if they are from known distribution families. The joint posterior in our case is somewhat atypical because of the selection problem. It may appear that the full conditional distributions of α_i and β_i as well as those for P_{it}^{*} are not from known distribution families if we just inspect them. However, on closer inspection, it turns out that they can be written as normal distribution.

In order to see this, it is useful to apply a simple trick of converting the joint density of d_{it}[19] into the product of conditional density of $(b_{it}|P_{it}^{*})$ and the marginal density of P_{it}^{*} when writing the full conditional density for α_i. Similarly, we write this joint density as the product of the conditional density of $(P_{it}^{*}|b_{it})$ and the marginal density of b_{it} when writing the full conditional density for β_i.

For this purpose, it is useful to note that if

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{12} & \Gamma_{22} \end{pmatrix}\right] \tag{B-1}$$

then

$$\theta_1|\theta_2 \sim N\left[\mu_1 - V_{11}^{-1}V_{12}\left(\theta_2-\mu_2\right), V_{11}^{-1}\right] \text{ where } V = \begin{pmatrix} V_{11} & V_{12} \\ V_{12} & V_{22} \end{pmatrix} = \Gamma^{-1} \tag{B-2}$$

[19]The topmost line in equation A-10.

Note also that

$$\Sigma^{-1} = \begin{pmatrix} \frac{1}{\sigma^2(1-\rho^2)} & -\frac{\rho}{\sigma(1-\rho^2)} \\ -\frac{\rho}{\sigma(1-\rho^2)} & \frac{1}{(1-\rho^2)} \end{pmatrix} \tag{B-3}$$

Thus

$$b_{it}|P^*_{it} \sim N\left[X_{it}\alpha_i + \sigma\rho\left(P^*_{it} - Y_{it}\beta_i\right), \sigma^2\left(1-\rho^2\right)\right] \tag{B-4}$$

and

$$P^*_{it}|b_{it} \sim N\left[Y_{it}\beta_i + \frac{\rho}{\sigma}\left(b_{it} - X_{it}\alpha_i\right), 1-\rho^2\right] \tag{B-5}$$

Thus, we can write the density of d_{it} (i.e. the joint density of b_{it} and P^*_{it}) in the first line of A-10 in the following two ways

$$f\left(d_{it}|\cdot\right) = f\left(b_{it}|P^*_{it},\cdot\right) f\left(P^*_{it}|\cdot\right) \tag{B-6}$$

and

$$f\left(d_{it}|\cdot\right) = f\left(P^*_{it}|b_{it},\cdot\right) f\left(b_{it}|\cdot\right) \tag{B-7}$$

Then, the full conditional distributions of α_i and β_i start looking like the familiar normal distribution kernels. Thus, we have the following full conditional distributions for the parameters

1. $\alpha_i|\cdot$

   In order to derive the full conditional distribution of α_i, we first rewrite the joint posterior density in A-10 using B-4 and B-6 and then select all the terms that contain α_i. We can ignore the other terms since they affect only the proportionality constant and not the kernel of the density.

$$f\left(\alpha_i|\cdot\right) \propto \begin{array}{c} \prod_{t=1}^{T_i}\left[\exp\left(-\frac{1}{2\sigma^2(1-\rho^2)}\left(b_{it} - X_{it}\alpha_i - \sigma\rho\left(P^*_{it} - Y_{it}\beta_i\right)\right)^2\right)\right]^{P_{it}} \\ \exp\left(-\frac{1}{2}\left(\alpha_i - M_\alpha z_i\right)' V_\alpha^{-1}\left(\alpha_i - M_\alpha z_i\right)\right) \end{array} \tag{B-8}$$

   By rearranging terms in the above expression, we can show that

$$\begin{array}{c} \alpha_i|\cdot \sim N\left[\tilde{\alpha}_i, \left(\tilde{X}_i'\tilde{X}_i + V_\alpha^{-1}\right)^{-1}\right] \\ \tilde{\alpha}_i = \left(\tilde{X}_i'\tilde{X}_i + V_\alpha^{-1}\right)^{-1}\left(\tilde{X}_i'\tilde{b}_i + V_\alpha^{-1}M_\alpha z_i\right) \end{array} \tag{B-9}$$

   where $\tilde{b}_i$ is the vector obtained by stacking $\frac{\left[b_{it}-\sigma\rho\left(P^*_{it}-Y_{it}\beta_i\right)\right]}{\sqrt{\sigma^2(1-\rho^2)}}$ for all t for which $P_{it} = 1$, and $\tilde{X}_i$ is the matrix obtained by stacking $\frac{X_{it}}{\sqrt{\sigma^2(1-\rho^2)}}$ for all t for which $P_{it} = 1$.

2. $\beta_i|\cdot$

The derivation of the full conditional distribution of β_i is similar to that for α_i. However, there is one significant difference. In the case of α_i, the full conditional was affected only by those observations where $P_{it} = 1$. In the case of the full conditional distribution of β_i, observations where $P_it = 0$ also enter, but differently from those where $P_{it} = 0$. Rewriting the joint posterior density in A-10 using B-5 and B-7 and selecting the terms that contain β_i gives us the following full conditional density

$$f\left(\beta_i|\cdot\right) \propto \quad \prod_{t=1}^{T_i} \left\{ \begin{array}{c} \left[\exp\left(-\frac{1}{2(1-\rho^2)}\left(P_{it}^* - Y_{it}\beta_i - \frac{\rho}{\sigma}\left(b_{it} - X_{it}\alpha_i\right)\right)^2\right)\right]^{P_{it}} \\ \left[\exp\left(-\frac{1}{2}\left(P_{it}^* - Y_{it}\beta_i\right)\right)\right]^{1-P_{it}} \\ \exp\left(-\frac{1}{2}\left(\beta_i - M_\beta z_i\right)' V_\beta^{-1}\left(\beta_i - M_\beta z_i\right)\right) \end{array} \right\} \tag{B-10}$$

The full conditional distribution of β_i can be shown to be

$$\begin{array}{c} \beta_i|\cdot \sim N\left[\tilde{\beta}_i, \left(\tilde{Y}_i'\tilde{Y}_i + \ddot{Y}_i'\ddot{Y}_i + V_\beta^{-1}\right)^{-1}\right] \\ \tilde{\beta}_i = \left(\tilde{Y}_i'\tilde{Y}_i + \ddot{Y}_i'\ddot{Y}_i + V_\beta^{-1}\right)^{-1}\left(\tilde{Y}_i'\tilde{P}_i^* + \ddot{Y}_i'\ddot{P}_i^* + V_\beta^{-1}M_\beta z_i\right) \end{array} \tag{B-11}$$

where $\tilde{P}_i^*$ is the vector obtained by stacking $\frac{\left[P_{it}^* - \frac{\rho}{\sigma}\left(b_{it} - X_{it}\alpha_i\right)\right]}{\sqrt{1-\rho^2}}$ for all t for which $P_{it} = 1$, $\tilde{Y}_i$ is the matrix obtained by stacking $\frac{Y_{it}}{\sqrt{1-\rho^2}}$ for all t for which $P_{it} = 1$, $\ddot{P}_{it}^*$ is the vector of stacked P_{it}^* for all t for which $P_{it} = 0$ and $\ddot{Y}_{it}$ is the matrix obtained by stacking Y_{it} for all t for which $P_{it} = 0$.

3. $m_\alpha = \text{vec}\left(M_\alpha\right)|\cdot$ and $m_\beta = \text{vec}\left(M_\beta\right)|\cdot$

$$\begin{array}{c} m_\alpha|\cdot \sim n\left[\tilde{m}_\alpha, \left(Z'Z + A\right)^{-1}\right] \\ \tilde{m}_\alpha = \text{vec}\left(\tilde{M}_\alpha\right) \text{ where } \tilde{M}_\alpha = \left(Z'Z + A\right)^{-1}\left(z'\alpha + A\bar{M}_\alpha\right) \end{array} \tag{B-12}$$

Z is the matrix formed by stacking z_i for all i, α is the matrix formed by stacking all α_i and $\bar{M}_\alpha$ is formed by stacking $\bar{\alpha}$ a column at a time.

Similarly,

$$\begin{array}{c} m_\beta|\cdot \sim N\left[\tilde{m}_\beta, \left(Z'Z + B\right)^{-1}\right] \\ \tilde{m}_\beta = \text{vec}\left(\tilde{M}_\beta\right) \text{ where } \tilde{M}_\beta = \left(Z'Z + B\right)^{-1}\left(Z'\beta + B\bar{M}_\beta\right) \end{array} \tag{B-13}$$

Z is the matrix formed by stacking z_i for all i, β is the matrix formed by stacking all β_i and $\bar{M}_\beta$ is formed by stacking $\bar{\beta}$ a column at a time.

4. $V_\alpha|\cdot$ and $V_\beta|\cdot$

$$\begin{aligned} V_\alpha|\cdot &\sim \text{Inverse Wishart}\left(\mu_\alpha + N, S_\alpha + E'_\alpha E_\alpha\right) \\ E_\alpha &= \alpha - Z\tilde{M}_\alpha + \left(\tilde{M}_\alpha - \bar{M}_\alpha\right)^{-1} A \left(\tilde{M}_\alpha - \bar{M}_\alpha\right) \end{aligned} \tag{B-14}$$

$\tilde{M}_\alpha$ and $\bar{M}_\alpha$ have already been defined in B-12 and N is the total number of consumers.

Similarly,

$$\begin{aligned} V_\beta|\cdot &\sim \text{Inverse Wishart}\left(\mu_\beta + N, S_\beta + E'_\beta E_\beta\right) \\ E_\beta &= \beta - Z\tilde{M}_\beta + \left(\tilde{M}_\beta - \bar{M}_\beta\right)^{-1} B \left(\tilde{M}_\beta - \bar{M}_\beta\right) \end{aligned} \tag{B-15}$$

5. $P^*_{it}|\cdot$

Using data augmentation (Tanner and Wong 1987), we can treat the vector of P^*_{it} as parameters and make draws for them from their full conditional distribution like in the case of the other parameters Note that when $P_{it} = 1$, P^*_{it} is drawn from a left-truncated normal distribution, while it is drawn from a right-truncated normal distribution when $P_{it} = 0$. Where our specific formulation differs from the standard binomial probit model is that the since we have a bivariate distribution of the errors, there are further restrictions placed on P^*_{it}.

$$P^*_{it}|\cdot \sim \begin{array}{ll} N\left[Y_{it}\beta_i + \frac{\rho}{\sigma}\left(b_{it} - X_{it}\alpha_i\right), 1 - \rho^2\right] 1\left(P^*_{it} > 0\right) & \text{when } P_{it} = 1 \\ N\left[Y_{it}\beta_i, 1\right] 1\left(P^*_{it} < 0\right) & \text{when } P_{it} = 0 \end{array} \tag{B-16}$$

6. $\sigma^2|\cdot$ and $\rho|\cdot$

The full conditional distributions of these two parameters are not from known distribution families and hence we cannot draw from them directly. We use the Random Walk Metropolis Hastings algorithm (Chib and Greenberg 1995) to make draws from these distributions. We use normal candidate densities for this purpose, using standard errors from likelihood estimates of these parameters and a scaling/tuning parameter that was set by maximizing the relative numerical efficiency to fix the variances of these candidate densities. More details are available from the authors on request.

Table 1: Summary Statistics - Gambling Behavior

Variable	Median	Mean	SD
Bet Amount ($)	131	448	1367
Previous Win ($)	20	38	319
Total Cum Win ($)	-46	-98	511
Cum Play Time (Min)	72	115	131
Total Play Time (Min)	124	170	165
Comp ($)	0	9	36

Table 2: Comparison of Gambling Behavior: Random and Frequent Visit Samples

Variable	Random Sample			Frequent Sample		
	Median	Mean	SD	Median	Mean	SD
Number of trips	2	7.8	17.0	24	32.5	24.3
Plays per trip	3	3.5	2.1	8	8.4	4.4
Total days in casino	2	8.3	20.5	27	37.4	30.0
Total Amount Bet in Trip ($)	131	448	1367	449	1508	4284
Total Amount Won in Trip ($)	-46	-98	511	-104	-182	1220
Total Play Time in Trip (Mins)	72	115	131	169	228	213
Comps per trip ($)	0	9	36	0	10	142

Table 3: Gamblers' Demographic Information

Variable	Random Sample	Frequent Sample
Age (mean)	61.1	57.2
Age (sd)	13.3	13.7
% Male	45.9	42.1
% Western European	64.6	60.3
% Hispanic	10.0	17.8
% Mediterranean	4.6	3.4
% Eastern European	3.7	2.3
% Other	17.1	16.2
% Prefer Slots to Games	85.9	66.2

Table 4: Individual-level Parameters: within-trip model

Parameter	Mean	Std. Dev.	5th pcntl.	Median	95th pcntl.
Bet sub-model					
Intercept	3.6839	0.1014	3.5308	3.6825	3.8546
LastBet	-0.0624	0.0115	-0.0794	-0.0627	-0.0430
LastWin	-0.0253	0.0233	-0.0607	-0.0253	0.0129
$LastWin^2$	0.0049	0.0025	0.0019	0.0049	0.0076
LastLoss	0.2572	0.0567	0.1606	0.2580	0.3454
$LastLoss^2$	-0.0026	0.0028	-0.0054	-0.0027	0.0000
CumWin	0.0260	0.0751	-0.0908	0.0244	0.1599
$CumWin^2$	-0.0021	0.0050	-0.0104	-0.0020	0.0055
CumLoss	-0.2635	0.2006	-0.5672	-0.2665	0.0784
$CumLoss^2$	0.0155	0.0206	-0.0177	0.0158	0.0468
Comps	0.1253	0.0084	0.1133	0.1255	0.1362
Play sub-model					
Intercept	0.9415	0.0225	0.9140	0.9388	0.9817
LastBet	-0.0149	0.0101	-0.0285	-0.0160	0.0046
LastWin	-0.1077	0.0031	-0.1117	-0.1080	-0.1022
$LastWin^2$	0.0163	0.0048	0.0091	0.0162	0.0236
LastLoss	-0.0988	0.0053	-0.1058	-0.0994	-0.0895
$LastLoss^2$	0.0221	0.0050	0.0140	0.0224	0.0278
CumWin	0.1262	0.0028	0.1232	0.1261	0.1302
$CumWin^2$	-0.0130	0.0049	-0.0200	-0.0133	-0.0053
CumLoss	0.0945	0.0150	0.0747	0.0926	0.1224
$CumLoss^2$	-0.0235	0.0071	-0.0332	-0.0241	-0.0077
Comps	0.0689	0.0079	0.0597	0.0682	0.0814
TimeSpent	-0.4211	0.0222	-0.4487	-0.4238	-0.3808

Table 5: Elasticities: within trip model

Variable	Bet Amount	Play Probability	Total
LastBet	-0.2529	-0.0297	-0.2825
LastWin	-0.0061	-0.0142	-0.0203
LastLoss	0.1643	-0.0310	0.1332
CumWin	0.0105	0.0265	0.0370
CumLoss	-0.3095	0.0645	-0.2452
Comps	0.1079	0.0287	0.1367
TimeSpent		-0.2357	

Table 6: Heterogeneity in individual-level parameters: within-trip model

Parameter	Intercept	Male	Age	Caucasian	Hispanic	Tablepref
Bet sub-model						
$Intercept$	3.7172**	-0.0144**	-3.72E-04*	-0.0045	-0.0461**	-0.0003
$LastBet$	-0.0580**	-0.0007	-7.66E-05**	-0.0002	-0.0048**	0.0017**
$LastWin$	-0.0231**	-0.0010	6.59E-07	-0.0016	-0.0037*	-0.0024
$LastWin^2$	0.0045**	0.0001	5.01E-06	0.0001	0.0004*	-1.78E-05
$LastLoss$	0.2372**	0.0070**	2.65E-04**	0.0020	0.0247**	-0.0015
$LastLoss^2$	-0.0027**	-5.19E-05	-7.87E-07	0.0002	-0.0002	2.21E-04
$CumWin$	0.0505**	-0.0100**	-2.93E-04**	-0.0026	-0.0329**	3.08E-04
$CumWin^2$	-0.0033**	0.0008**	1.21E-05	0.0001	0.0019**	-3.19E-04
$CumLoss$	-0.1948**	-0.0278**	-8.04E-04**	-0.0087	-0.0921**	-5.55E-04
$CumLoss^2$	0.0091**	0.0034**	6.08E-05	0.0010	0.0094**	-1.15E-04
$Comps$	0.1216**	0.0002	6.25E-05**	0.0005	0.0031**	1.26E-04
Play sub-model						
$Intercept$	0.9450**	-0.0049**	3.19E-05	-0.0017	-0.0067**	-0.0038**
$LastBet$	-0.0130**	-0.0015**	-8.61E-06	-0.0002	-0.0034**	-4.05E-04
$LastWin$	-0.1071**	-0.0006**	1.29E-06	-0.0002	-0.0008**	-5.93E-04**
$LastWin^2$	0.0163**	0.0001	3.66E-06	-0.0002	-0.0004	-2.62E-04
$LastLoss$	-0.0979**	-0.0011**	5.87E-06	-0.0003	-0.0018**	-0.0012**
$LastLoss^2$	0.0216**	0.0002	6.57E-06	3.79E-05	0.0003	4.29E-04
$CumWin$	0.1264**	-0.0005**	7.25E-06	-0.0001	-0.0003	-5.84E-04**
$CumWin^2$	-0.0124**	-0.0001	-6.12E-06	-0.0002	-0.0006	-7.60E-04**
$CumLoss$	0.0973**	-0.0030**	1.17E-05	-0.0012	-0.0052**	-0.0023**
$CumLoss^2$	-0.0210**	-0.0007*	-4.00E-05**	-0.0002	-0.0016**	-8.78E-05
$Comps$	0.0690**	-0.0014**	2.46E-05	-0.0004	-0.0014**	-7.02E-04
$TimeSpent$	-0.4160**	-0.0046**	-9.36E-06	-0.0014	-0.0069**	-0.0034**

Table 7: Long Run Comp Elasticities: Effect of Addiction

	Total Amount Bet	Number of Plays In Trip
Estimated Parameter Values	0.0781	0.0826
Non-Addicted Individuals	0.0752	0.0812
Addicted Individuals	0.1843	0.1595

Table 8: Individual-level Parameters: across-trip model (Random Sample)

Parameter	Mean	Std. Dev.	5th pcntl.	Median	95th pcntl.
$Intercept$	1.2234	0.6214	0.3043	1.4132	2.7534
Bet_{it}	-0.0453	0.0815	-0.1731	-0.0462	0.0499
Win_{it}	-0.7639	1.5932	-3.3074	-0.7170	2.0482
Win_{it}^2	-0.0312	0.0637	-0.1132	-0.0331	0.0342
$Loss_{it}$	0.3985	0.1612	0.1176	0.4173	0.5888
$Loss_{it}^2$	-0.1630	0.0853	-0.2839	-0.1653	-0.0246
$Comps_{it}$	0.0004	0.0212	-0.0236	-0.0007	0.0174
$TimeSpent_{it}$	-0.3221	0.2131	-0.6497	-0.3132	-0.0053

Table 9: Individual-level Parameters: across-trip model (Frequent Sample)

Parameter	Mean	Std. Dev.	5th pcntl.	Median	95th pcntl.
$Intercept$	2.6057	0.4909	1.7066	2.6637	3.3253
Bet_{it}	-0.0481	0.0648	-0.1614	-0.0470	0.0528
Win_{it}	-0.7232	2.4393	-6.1222	-0.6592	3.0564
Win_{it}^2	-0.0381	0.0525	-0.1077	-0.0391	0.0351
$Loss_{it}$	0.4293	0.1708	0.1288	0.4432	0.6909
$Loss_{it}^2$	-0.1571	0.0802	-0.2751	-0.1636	-0.0228
$Comps_{it}$	0.0006	0.0151	-0.0178	0.0000	0.0192
$TimeSpent_{it}$	-0.2217	0.1847	-0.5277	-0.2192	0.0709

Figure 1: Scatterplot of individual level *LastBet* and *Comps* coefficients play submodel

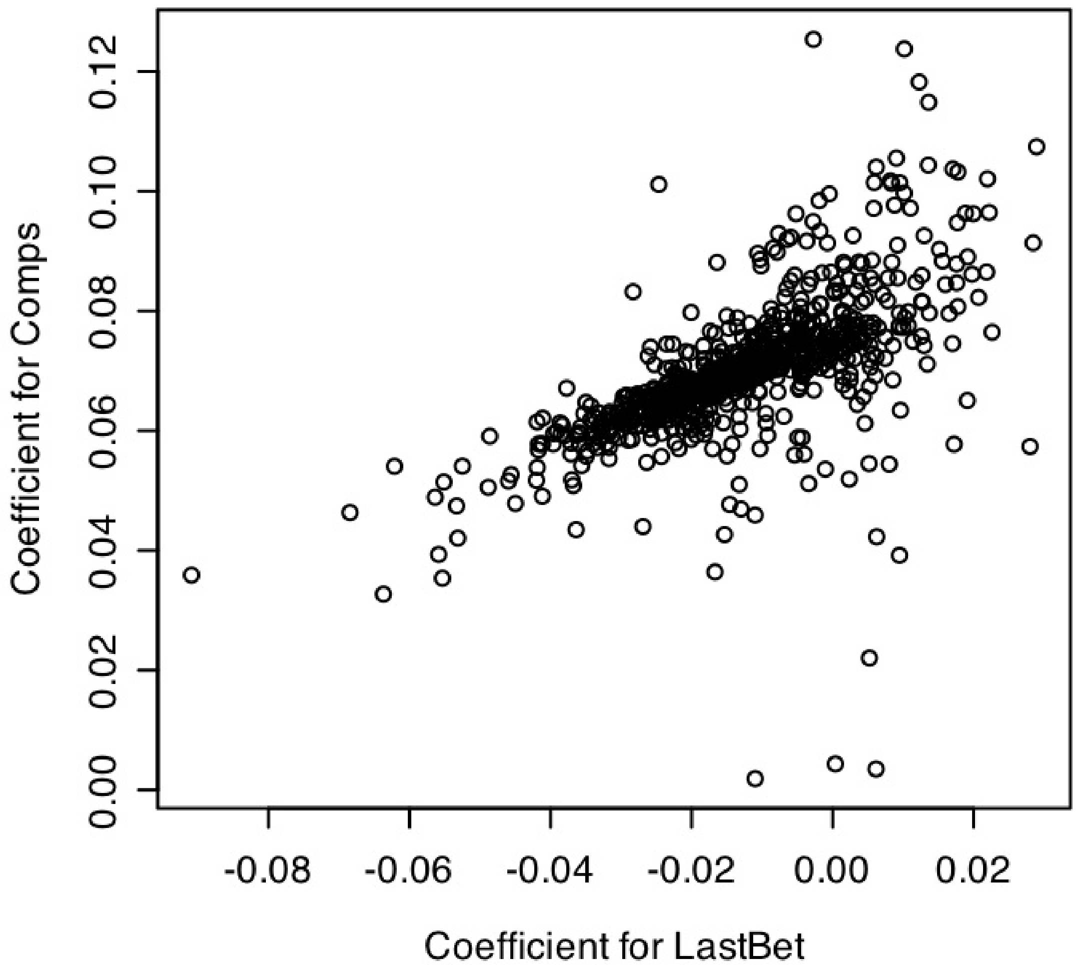

Figure 2: Scatterplot of individual level *LastBet* coefficients and *Intercepts* for the bet sub-model

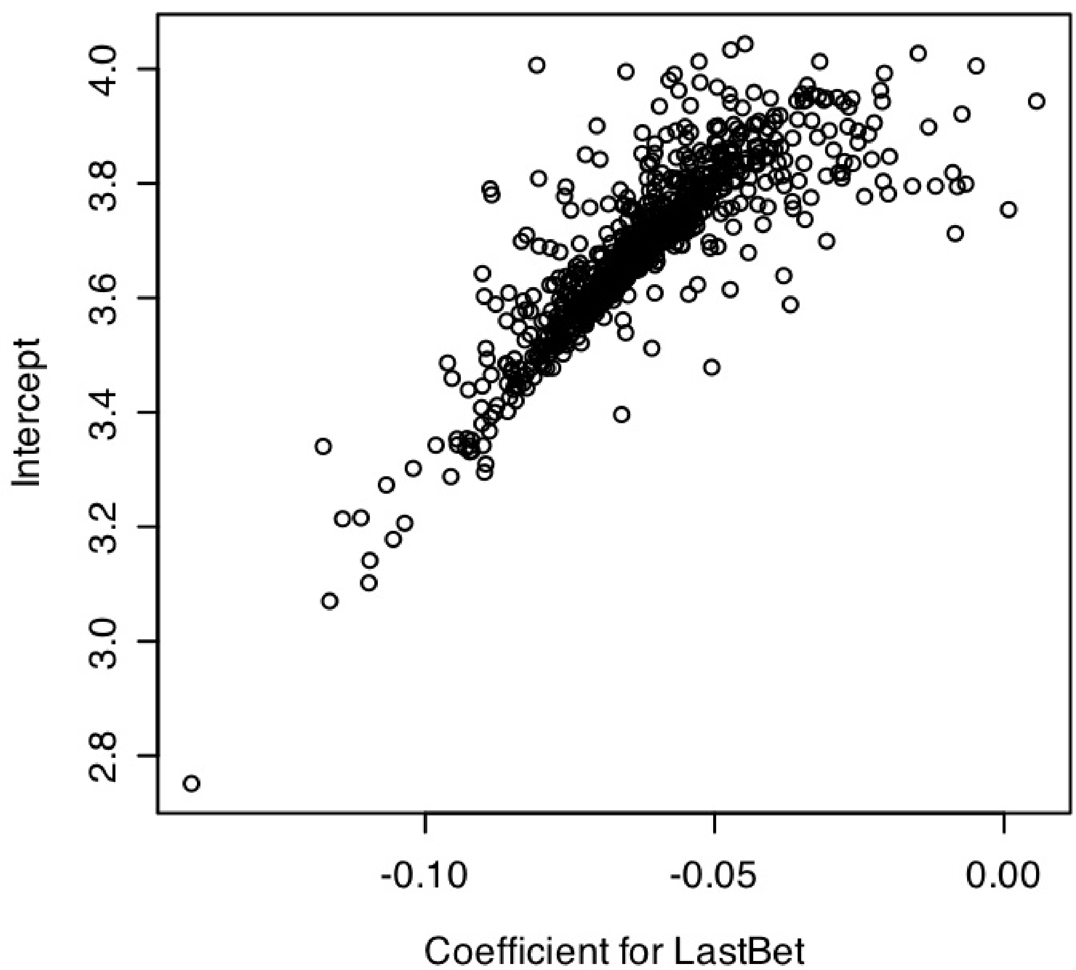

Figure 3: Scatterplot of individual level *LastBet* coefficients and *Intercepts* for the play sub-model

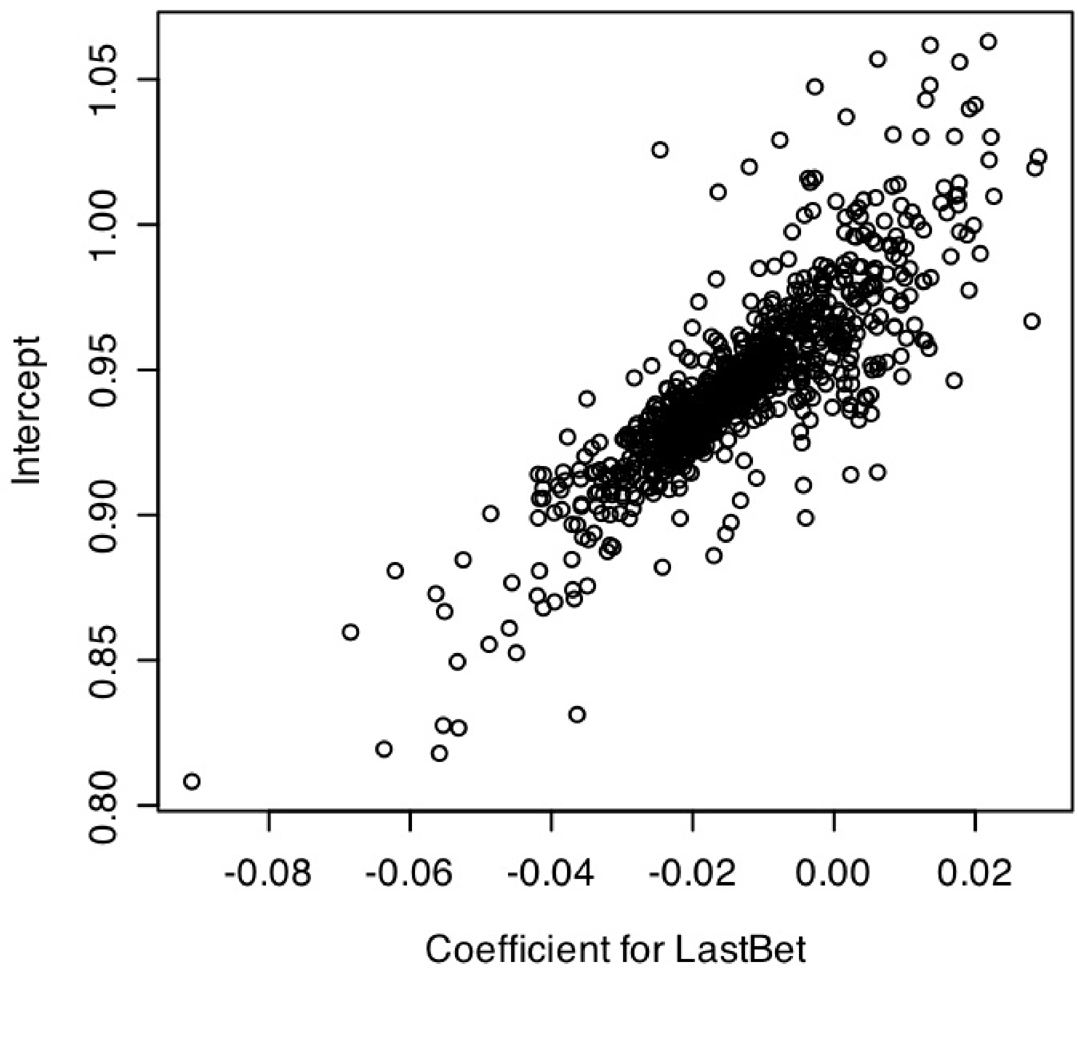

Figure 4: Scatterplot of individual level *LastBet* coefficients for within and across trip models (frequent sample)