

2016

Essays on Optimal Efforts, Costs and Rewards

Soumyadip Roy
Iowa State University

Follow this and additional works at: <https://lib.dr.iastate.edu/etd>

 Part of the [Economics Commons](#)

Recommended Citation

Roy, Soumyadip, "Essays on Optimal Efforts, Costs and Rewards" (2016). *Graduate Theses and Dissertations*. 15803.
<https://lib.dr.iastate.edu/etd/15803>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Essays on optimal efforts, costs and rewards

by

Soumyadip Roy

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Economics

Program of Study Committee:

GianCarlo Moschini, Co-Major Professor

Peter F Orazem, Co-Major Professor

John Schroeter

Sunanda Roy

Gray Calhoun

Iowa State University

Ames, Iowa

2016

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iv
CHAPTER 1. STRATEGIES AND OUTCOMES IN DYNAMIC ASYMMETRIC GAMES: THE CASE OF SOCCER	
1. Abstract	1
2. Introduction	1
3. Contest Theory in Economics: A Brief Review	2
4. Contest Theory and Sports	8
5. Strategy in Soccer: An Empirical Study	13
6. A Strategic Model of Soccer	14
7. A Strategic Model of Soccer: Empirical Results	22
8. Skill Variable	26
9. Econometric Model and Results	28
10. Conclusions	31
11. References	44
CHAPTER 2. HEALTH, TYPES OF LEISURE AND TIME ALLOCATION: EVIDENCE FROM ATUS DATA	
1. Introduction	47
2. Theoretical Model	49
3. Econometric Model	54
4. Data	56
5. Results	60
6. Conclusions	68
7. References	84

CHAPTER 3 EFFECTS OF MINIMUM WAGE ON EMPLOYMENT,
EARNINGS AND WAGE BILL: EVIDENCE FROM THE RETAIL SECTOR
IN USA (1991 -2014)

1. Introduction	86
2. Literature Review and Motivation	87
3. Empirical Strategy	89
4. Data	99
5. Results	100
6. Conclusions.....	104
7. References	108

ACKNOWLEDGMENTS

I would like to thank my committee chairs, GianCarlo Moschini and Peter Orazem, for their expert guidance in writing this dissertation. I am also thankful to my committee members - John Schroeter , Sunanda Roy, and Gray Calhoun, for their guidance and support throughout the course of this research. In addition, I would also like to thank my friends, colleagues, the department faculty and staff for making my time at Iowa State University a wonderful experience. This dissertation would not be possible without the inspiration of my parents and the support of my wife.

CHAPTER 1

STRATEGIES AND OUTCOMES IN DYNAMIC ASYMMETRIC GAMES: THE CASE OF SOCCER

Abstract

This paper uses contest theory to derive a multi-stage model of soccer matches. Numerical simulations are used to derive several conjectures that are testable based on observable outcomes. The econometric analysis of these conjectures relies on an original large dataset that we have assembled, pertaining to all matches played in seven years of the Italian Serie A and five years of the Spanish La Liga. Our results strongly support the notion that the state of the match influences the equilibrium strategies adopted by the players. In particular, we find that the probability of observing a goal increases as the match gets closer to the end, conditional on the state of the game being a tie. The role of asymmetry in skill levels, fatigue, end of season effects are also considered.

1. Introduction

Within the framework of neo-classical economics, competition is believed to encourage allocation of scarce resources to their best use and is believed to increase efficiency. In many economic situations, competition can be represented in terms of so-called contests. These strategic settings include advertising and marketing by firms, patents races, rent-seeking by interested parties, litigation, military conflicts and sports. A considerable body of work in these areas has provided numerous theoretical insights. Empirical studies are less common. In part, this is a reflection of the difficulties of isolating the strategic implications of contests from the many other factors that impacts real-world data. Hence, economists have occasionally turned to sports data for empirical applications. The appeal of such data, in this setting, is because the nature of sporting competitions is such that they closely resemble standard contests: players perform under strictly laid down rules, and outcomes are readily observable. Hence, sporting competitions provide rich datasets that allow us to empirically test economic theories. In this paper, we develop and test some game-theoretic predictions, derived from a multistage contest, with real world data from soccer.

This paper is organized as follows. Section II provides an introduction to the theory of contests and reviews areas in which this theory has been broadly applied in economics. Section III describes applications of contest theory in sports. Section IV deals with strategic issues in soccer. Section V explains the model used for the numerical exercise and motivates the testable hypotheses. Section VI provides description of the data. Section VII describes how the Skill variable is derived. Section VIII describes the econometric model and discusses results. Section IX concludes. Tables and figures are at the end of the paper.

2. Contest Theory in Economics: A Brief Review

It is useful to start by formally introducing the notion of contests—Konrad (2009) provides a comprehensive overview. A contest is a game, involving n players (the contestants). Each contestant $i = 1, 2, \dots, n$ puts in a certain amount of effort and competes for a prize of size b . The vector of efforts $x = (x_1, x_2, \dots, x_n)$ determines the winner. The allocation of the prize among contestants can take different forms. However, in the simplest "winner-takes-all" type of contest, only one contestant wins the prize of value b , while everyone else gets zero. The valuation of the prize may differ among contestants. If contestant i 's valuation is denoted as $v_i(b)$, and the cost of effort for player i is denoted by $C_i(x_i)$, then contestant i 's payoff is written as:

$$\pi_i(x_1, x_2, \dots, x_n) = p_i(x_1, x_2, \dots, x_n) \cdot v_i(b) - C_i(x_i)$$

Where $p_i(x_1, x_2, \dots, x_n)$ denotes the "Contest Success Function", with $0 \leq p_i \leq 1$ and $\sum_{i=1}^n p_i = 1$.

2.1 Rent-Seeking Contests: When a market economy operates in the presence of a government, there are special privileges that can be earned by interested parties. The nature of rent that is relevant here arises from an artificial contraction of output and interested parties can earn returns in excess of their opportunity costs by expending effort in 'rent-seeking' activities as opposed to 'profit-seeking' activities, where only the latter is seen as productive from a societal point of view. Rents may emerge from shifts in demand and supply curves and the pursuit of such rents are equivalent, in principle, to 'profit-seeking' (Tollison 1982). However, rents that arise from an artificial contraction of output are not productive and any effort expended by interested parties to earn them are wasteful from the point of view of society. Having made this distinction, it is important to note that even these artificially created rents cease to exist in the long-run, i.e., tend to be dissipated by competition.

In the rent-seeking setup, a contest is a strategic game with $n \geq 2$ players and pure strategy spaces are described by the vector of efforts x and the set of payoff functions described above. The literature on this subject has looked at the existence and uniqueness of Nash equilibria as well as the

extent of rent dissipation, where rent dissipation D is defined as: $D = \frac{\sum_{i=1}^n x_i^*}{b}$. Here, the vector $x^* = (x_1^*, \dots, x_n^*)$ denotes Nash equilibrium efforts. The numerator here represents the total rent-seeking outlays in equilibrium and the denominator is the value of the contested rent. Before going into the economic significance of D , let us introduce the special case of Tullock's rent-seeking contest. This is among the most popular contest success functions and is the one that assumes that a contestant's probability of being the winner is a function of the ratio of his own effort and the sum of all contestants' efforts (Konrad 2009).

In the Tullock-contest, identical agents are risk-neutral and the cost of effort is assumed to be linear with unit marginal cost. Specifically, the payoff for individual i is written as:

$$\pi_i(x_1, x_2, \dots, x_n) = \frac{x_i^r}{\sum_{j=1}^n x_j^r} \cdot v_i(b) - x_i$$

where the parameter r determines the discriminatory power of the contest success function. With a high r , an effort level slightly higher than one's rival ensures a high probability of a win whereas with a low r , the difference in effort levels does not make much difference.

The beauty of this form of the contest success function is the fact that it is imperfectly discriminatory - the contestant with the highest effort level is not sure to win except the case when all other contestants expend zero effort. This is in contrast, for example, to the all-pay-auction scenario where the highest bidder always wins. This functional form may thus be used to study realistic scenarios like sports competitions where a key attraction is the fact that the underdog can beat the favorite. This contest success function also captures the importance of other factors such as strategies employed by teams/individuals, fatigue and environmental factors such as home advantage, familiarity with the playing conditions etc. by leaving room for the possibility of an uncertain outcome.

2.2 Rent-Dissipation: Over dissipation of rents ($D > 1$) is a concern primarily because it implies expenditure of wasteful resources greater than the value of the prize. Baye, Kovenock and De Vries

(1993) show that a symmetric Nash Equilibrium exists in Tullock-type contests with two players and $r > 2$ and prove that over-dissipation of rents do not occur in any Nash equilibrium. This property is actually one of the reasons why Tullock's version of the rent-seeking contest is appealing and popular. In the Tullock-type contest, rent dissipation increases as the number of players increase and rent is totally dissipated ($D = 1$) when the number of players is sufficiently large. It might be useful to mention here that the parameter r can alternatively be looked upon as the marginal return to lobbying outlays and it should be noted that D increases with an increase in r .

Consideration of asymmetry and uncertainty in contests are relevant to the current study. We do not discuss Risk Aversion in detail here, however, primarily because Risk-Neutrality with linear payoffs are standard assumptions in the context of contest theory. See Nitzan (1994) for details.

2.3 Asymmetry And Dynamics: Asymmetry in the rent-seeking contests may arise from heterogeneous valuations of the prize or heterogeneity in individual lobbying capabilities caused by different costs of effort for different players. This is similar to sports contests where participants may have heterogeneous valuations for the prize or different costs of effort. Hillman and Riley (1989) consider asymmetric valuation of the prize by rent-seekers. They show that such asymmetric valuations may act as a barrier to entry and the incumbent firm may have certain advantages in a monopoly rent situation. It is shown that free entry in a perfectly discriminating contest does not imply that all players are active participants in the contest if asymmetric prize valuations are present. In reality, the two agents with the highest valuations of the prize are the only ones who have incentive to actively fight the contest.

Leininger (1993) reinforces the point made by Hillman and Riley (1989) that asymmetric valuations decrease overall rent-seeking expenditures. This paper questions the assumptions made in the earlier literature regarding simultaneous move by both players and introduces dynamics into the rent-seeking game. He assumes that players can choose the time when they announce their effort levels in addition to choosing their actual effort levels. Asymmetry with regard to types of players here follows the setup in Hillman and Riley (1989) with asymmetric valuations. The players expend different effort levels and these effort levels determine the probabilities of winning. A bias in this case is brought out by a higher evaluation of the prize by a player (known as the 'stronger' player) and this increases a player's expected payoff in equilibrium in the simultaneous Cournot-Nash cases. For the Stackelberg case where sequential play takes place in Stage 2, the paper gives us an important result -- with players of different strength, there exists a unique subgame-perfect equilibrium in the two-stage game where the weaker player moves first and the stronger player moves second. After observing each

other's strengths (this is assumed to be common knowledge), the stronger player realizes that he does not need to 'flex his muscles' fully since the weaker player would not be able to match him and part of his efforts would be wasteful. The weaker player realizes that any opening bid that is too bold would face even bolder bidding by the stronger player and he would not be able to match the stronger player in such a contest. Thus, the only possible solution where deviation isn't beneficial for both parties is when the weaker player opens the game with a moderate bid and the stronger player responds with an even more moderate bid, yet getting a better expected share of the prize because his higher valuation increases his chances of winning. Both players prefer this situation to the one where they try to outbid each other. This model provides important nuances for the asymmetric two-player contest which we use in our model for the game of soccer described below.

Leininger and Yang (1994) re-examine the same problem in a dynamic context and allow for finitely or infinitely repeated alternative bids by players contesting for the rent. This paper shows that the infinitely repeated model is more appropriate for the context of rent-seeking where it is not clear who the last mover is. It is shown that wasteful competition is reduced by this form of sequential bidding and 'implicit collusion' results. Rent-dissipation is even smaller than the static versions of the game studied before. Tit-for-tat strategies form an SPE (Subgame Perfect Equilibrium) for the game they analyze. The tit-for-tat strategy is characterized by threats and counter-threats and this restrains competitors from entering into escalatory behavior. Leininger and Yang (1994) term this to be 'implicit collusion' and such collusion emerges from the dynamic nature of the game. This sense of 'implicit collusion' is in accordance to the idea in the Leininger (1993) paper where the players do not go at each other with full strength and mutually benefit from moderate bids. As we will see in the empirical application, this situation may arise in soccer where skill levels of teams are similar and their position in the league table is such that they may mutually benefit from collusion.

2.4 Commitment And Order Of Play: Dixit (1987) analyzes strategic behavior in contests. He looks at commitment in a contest between a favorite team and an underdog. He shows that under both the logit (a special case of which is the Tullock-contest) and probit specifications of the contest success functions, the favorite has an incentive to overcommit to effort, a result which does not arise in the perfectly symmetric two-player game. It is also proved that for both logit and probit specifications, the player favored to win has a strategic incentive to over-commit and the underdog to under-commit. In the rent-seeking scenario, this result may be interpreted as follows: the more efficient user will put in more effort

if they are given first access to a bureaucrat (who, say, has the power to sign one party into a lucrative contract) as compared to when simultaneous access is given to both players.

Baik and Shogren (1992) use Dixit's model to show that in an SPE, the favorite never moves first. They also prove that contests with asymmetric players typically lead to underinvestment of efforts and lower social costs than in Nash equilibria with symmetric players.

2.5 Some Economic Applications: Beauty Contests, Quantitative Restrictions And Corruption:

'Beauty contests' refer to contests where rent-seekers spend effort to win rents which arise either from restrictive trade policies, development policy of the government or industry regulation. An example is when firms spend resources in lobbying efforts to try and obtain import licenses in a country with import restrictions. Krueger (1974) looks at this issue in India and Turkey and argues that import licenses impose costs on society which are in addition to the traditional triangular dead weight loss due to rent-seeking by interested parties. Policy decisions are tricky here, because if the government restricts entry then it might be seen as favoring certain 'special groups' whereas if it allows free entry then it increase wasteful rent-seeking activity. The suspicion on the free-market, particularly in developing countries, often stems from this issue. If the free-market is suspect of favoring firms with greater resources and unequal distribution of income is seen as a by-product of this system, then the political class agrees to intervene more into the operations of the free market and this creates artificial barriers. Krueger (1974) points out that this usually leads to the development of a political 'vicious cycle' whereby restriction leads to rent-seeking and then further intervention from the government results.

Hazlett and Michaels (1993) analyze costs incurred due to the Federal Communications Commission's (FCC) cellular telephone license lotteries held from 1986 to 1989. Two licenses were awarded per market, one to a telephone company (local exchange carrier) and the other to a 'non-wireline' company. Since the requirements for obtaining a license were impossible to meet for more applicants, after the announcement of the application method, a large industry started with some firms known as 'application mills' developing standard application forms with appropriate engineering specifications. For the empirical estimation, the price of the licenses is taken as a proxy for their value because these licenses were often bought and resold. Dissipation was not perfect. While there were no discernable barriers to entry, incumbents may have had some advantage in reducing entry thereby leading to imperfect rent-dissipation. However, the FCC's decision to switch to auctions for allocation of spectrum later on reduced rent-seeking activity by a large extent.

2.6 Education Filters: Arrow's (1972) paper on education filters assumes that higher education does not contribute to superior economic performance but acts as a screening device by sorting out individuals of differing abilities. This tournament aspect of education stems from recognizing education as a system that enables appropriate matches of employers with candidates of suitable ability, where only relative abilities matter. In this regard, some of the effort made towards education may be wasteful and the nature of the tournament may be of the rent-seeking type.

Amegashie and Wu (2004) use this setup of two types of students - high and low who get admitted to two types of schools - high quality and low quality and use an all-pay auction setup. Using the school admission system in China they show that misallocation of students to schools is possible, with high-ability students choosing low-quality schools and vice versa.

Fu (2006) looks at affirmative action in education and shows that there is no tradeoff between academic quality and affirmative action policies. An affirmative action policy may endogenously emerge in equilibrium even when the sole objective of universities is to maximize academic quality.

2.7 Nested Contests And Military Conflicts: Military conflicts closely resemble contests. Countries produce weapons by spending resources that are costly and such resources have opportunity costs in terms of consumption goods. Some of these resources are wasted in the event of no-war and thus the structure of military conflicts closely resembles that of contests. As Konrad (2009) notes, empirical studies of military conflicts show that a number of factors other than the size of the army might be responsible for winning a military battle, such as superior technology, better war strategy, leadership, morale and possibly luck. Also, once the war is won, sharing rules among group members becomes important. And inter-group contest may be followed by an intra-group one. This situation may also arise in team sports.

2.8 Group Contests: Olson (1965), in the book "The Logic of Collective Action," discusses group contests. The central argument is that the theory of rational self-interested economic agents coming together to fight for the maximization of a group cause is inherently inconsistent. He goes so far as to say that members of a big group would not act to advance the group's common objectives unless they are forced to do so or if they are promised of some payment in addition to the group winnings in case they win. He asserted that the free-rider problem in large groups is so acute that, in reality, large groups exert less aggregate effort than small groups. Corchon (2007) discusses this in detail.

3. Contest Theory and Sports

The economic theory of contests has been tested with real-world empirical data. There is vast empirical literature on a variety of areas including R&D contests, political campaigning, committee brining, vote buying, military conflict and sports. Because this paper uses sports data, it may be helpful to look more closely at sports applications of contest theory.

3.1 Strategic Effects: As discussed above, asymmetry may arise from differences in valuations of the prize by the two contestants, as well as in the underdog-favorite setting studies by Dixit (1987). However, asymmetries may arise due to differences in abilities as well. As Szymanski (2003) notes, in a symmetric contest, there is no trade-off between winning effort, average effort and the variance of effort. However, an important aspect of the design of an asymmetric sporting contest between two players is whether the organizers want to maximize the winning effort (for example breaking a world record) or average effort. If incentives for a win are increased or a second prize is introduced, then weaker contestants might be encouraged to put in more effort. The role of the second prize, however, becomes important in asymmetric three-person contests where two players are relatively weak and one is strong. In the absence of a second prize, the strong contender does not exert full effort as soon as he realizes the strength of the other two contestants. However, with a second prize, the weaker contestants fight for it and this makes sure that the stronger contestant cannot take it too easy.

Most sporting competitions are sequential in nature. Teams play against each other for a number of games and the team that wins the maximum number of matches wins the competition. This is true in elimination tournaments, best of n-matches tournaments as well as leagues. The natural way to solve a sequential game of such kind is to look for Subgame Perfect Nash Equilibria (SPNE). The sequential nature of these contests creates a strategic element to the game. Teams adjust their effort levels according to the state of the match, their position in the league/ tournament as well as the quality of their opposition.

Economists have looked at the strategic aspect of sports at length. In Basketball, there is a concept of momentum known as 'hot hand' or 'streak shooting'. It is believed that the probability of hitting a shot is higher following a hit than following a miss. Psychological momentum, however does not explicitly account for the choice of effort but it does modify the probability of winning in a particular round according to the outcomes of the previous rounds. The main issue here is - if sequential rounds of play are independent from each other or not.

Ferrall and Smith (1999) find negligible strategic effects in baseball, basketball and hockey. They make the bold claim that teams play as well as possible in each match regardless of the situation of the match or the position of the team in the tournament. Their theory is consistent with players being robotic effort maximizers who play without passion or possibly, the same amount of passion in every minute of every game. Teams are exactly as good as they are on paper.

Leach (2005) argues that tennis provides a better test of incentive responses in a tournament setting. This is because tennis is an individual sport and the issues regarding sharing of the prize and free-rider problems do not arise in it. Challenges with respect to identification of the reward structure are as follows - while the size of the final prize of the game is known, it is difficult to know the value of winning a particular game in a particular set. Players are assumed to be risk-neutral, who maximize their expected net payoffs in the match, which is common knowledge. The empirical analysis, using Association of tennis professionals (ATP) data from the 1998 and 1999 seasons show that there exist significant strategic effects. Players change their effort levels according to the state of the match and the reward structure at a particular stage-game.

Malueg and Yates (2010) reject the independent probability model using ATP data from 2002 to 2006. They consider best-of-three tennis matches between players of similar ability and derive strong results in a simple setup. Their argument is: In best of three models, the winner of the first rounds exerts more effort in the second round than the loser of the first round. If the winner of the first round loses the second round, then the situation is symmetric as in the first round, in the third round. Thus the assumption that the result of the first round does not have any impact on the game play in the second round is incorrect, indicating towards the presence of strategic effects in tennis matches.

Walker and Wooders (2001) mention that professional players not only know how to play the sport, but they know how to play it well. Thus experiments where subjects are asked to play a game may not sufficiently capture the complex nature of sporting competitions because the participants in experiments are not as skilled as professional players. While mentioning that experiments might still be superior to empirical tests of the minmax theorem, use of which has become standard in sports situations, they analyze 2 X 2 games as a theoretical model of serves and its relation to the winning of points in a tennis match. The data is constructed from videotapes of ten tennis matches between top-ranked players. The paper fails to reject the hypothesis that the probability of serving to the right of the receiving player is the same as that to the left of the receiving player.

As Szymanski (2003) notes, most of the literature on the economics of team sports has focused on normative issues such as gate sharing rules, dependence of winning percentages on the wage bill, dominance of certain teams over a period of time indicating towards reduced competitive balance etc. In this paper, however, we focus on strategic issues that arise as the sport contests are played. Before going into our current project in detail, it may be useful to summarize two key results in contest theory that hold up in empirical analyses of contests as well as experiments on contests:

(1) In contests, increased efforts by one participant(s) is a **negative externality** for the other player(s). This is related to the heterogeneity of contestants - as heterogeneity among contestants increases, aggregate effort decreases and a particular contestant can increase her probability of winning by putting in more effort.

(2) Contestants with **higher valuations of the prize expend higher efforts** and are more likely to win the prize. Baye, Kovenock and De Vries (1996) show that there exists a relationship between differences in the valuation of a prize and differences in cost of efforts. This implies that low cost-of-effort individuals provide higher effort and are more likely to win the contest.

3.2 Strategic Issues in Soccer:

The economics literature is divided on what causes teams to play more offensively or defensively in soccer. It may be useful to mention here that FIFA introduced the 3-points-for-a-win system in the group-stage of the 1994 FIFA World Cup, with the hope of increasing attacking play. This led to most countries changing into this system in their domestic leagues and abolishing the 2-points for a win system that was previously in place. However, economists are divided on their opinion regarding the effectiveness of this rule in achieving the desired goals.

The first theoretical model on the merits of the three-point-for-a-victory rule is due to Brocas and Carrillo (2004). The 'three-point-for-a-victory' rule was adopted to encourage more attacking play than the two-points for a win rule since a tie (draw) becomes less attractive if three points are awarded to the team that wins. The approach of the Brocas and Carrillo (2004) paper is typical of most of the papers written on this topic - it focuses on incentives for strategic allocation of effort between attack and defense and not on optimal total effort. While it is true that efforts are costly, it can be argued that it is the choice between attack and defense that is key in soccer. It is this choice that determines the winner. While this choice inherently depends on the amount of effort chosen, the players on the pitch position themselves in a way that increases their chances of scoring when they have the ball and reduces the opponents' chances of winning when they have the ball. By focusing on

the choice of attack versus defense, these class of papers do not compromise on the basic nature of contest theory where optimal allocation of costly efforts is the cornerstone of analysis; instead they go one step ahead in their analysis and look at behavior in strategic play in an uncertain world and analyze how teams choose to allocate their costly efforts between attack and defense. The assumption is that everybody always gives their best on the football field (free-riders can be identified easily by the manager and substituted).

The paper by Brocas and Carrilo (2004) builds a theoretical model and shows that raising the reward for a victory may induce teams to play more defensively in the first half instead of increasing the overall level of attacking play. Also, increasing the value for a victory may induce teams to be more offensive towards the end of a game conditional on the game finishing as a tie. Thus, the dynamic nature of a soccer match becomes critical while analyzing it and no sweeping statement such as - 3-points for a win induces more attacking play in soccer can be made. The optimal strategy for teams is crucially dependent on the current score of the match. This paper stresses on this point and uses the state of the game (defined as the number of goals a team is ahead or behind by) as an explanatory variable in the econometric analysis. Dilger and Geyer (2009) also look at the effects of the three-point rule and show that leading teams tend to play more defensively. Their results are in accordance with the predictions of Brocas and Carrillo (2004). At the heart of both these papers is the 'option value' argument, which states that - when the change in payoffs between a tie and a defeat is smaller than the absolute change in payoffs between a tie and a victory, teams have incentives to play safely towards the beginning of matches.

The focus on incentive structures and their impact on soccer strategies is also the theme of the Moschini (2010) paper. Moschini also looks at the effect of the 3-point-rule on attacking play and builds a game-theoretic model on the lines of Brocas and Carrillo (2004). This paper assumes complementarity in attacking choices for teams in terms of producing goals and proves that these choices are increasing in the number of points awarded for a victory in a one-shot game. This paper shows that if the parameterization in Brocas and Carrillo (2004) is used, then an increase in the number of points awarded for a win can actually decrease the equilibrium action choices in stage 1. Analytical results that apply in a general context are not applicable in the context of this paper since the super-modularity condition and the increasing differences condition might not hold in stage $t=1$. Thus an alternative parameterization is used for a simulated model and Markov Perfect Nash Equilibria (MPNE) for a two-stage game are computed for a large number of parameter values. The

simulation results show that the observable implications of the static model are robust and extend to the multi-stage setting for a fairly general parametric combination. The predictions of the game-theoretic model are then analyzed empirically, by using a dataset from 30 countries over 35 years. The key objective of the current paper is to extend this analysis in the multi-stage setting and look at how the probability of a goal being scored evolves as the game progresses.

Unintended consequences of changes in the reward system such as sabotage activities have also been given considerable attention in the literature. Lazear's (1989) theoretical model, which investigates incentive structures in the labor market and concludes that there exist incentives for agents to 'make opponents look bad' when workers are paid on relative performance, is tested by Corral et al (2010) using data from the Spanish league. They confirm Lazear's prediction by showing that *ceteris paribus* an increase in rewards for a victory leads to an increase in number of red cards as a whole compared to previous seasons. Red cards are issued when players commit fouls that are deliberately aimed at stopping the opponents' attack with an intention of sabotage. The paper also shows that the number of red cards increase towards the end of games when rewards for a win are increased. The degree of offensive play is also analyzed by Haugen (2008) who argues that although an increase in rewards leads to more offensive play on average, there is a trade-off as it leads to more competitive imbalance. Such unintended consequences may lead to matches being one-sided, thereby defeating the premise of the rule change which assumes that more goals and lesser number of drawn matches make the sport more attractive to audiences.

Santos (2014) makes a more recent contribution to this literature and derives teams' 'optimal strategies' by using Principal Component Analysis (PCA). The empirical analysis in this paper shows that teams on an average do not adopt 'optimal strategies'. The paper shows that teams should attack as much as possible in almost all situations. The defensive nature of play, which is often observed in competitive soccer matches, is not optimal. This may be due to coaches who are imperfect maximizers or they might have non-standard objective functions. Palomino et al. (2000) use a game-theoretic approach to a soccer match and test the theoretical predictions using data from England and Spain. Their paper looks at state variables such as the Goal difference, Skills of different teams, Home advantage etc. and the theoretical predictions are confirmed by the empirical analysis. Our study is quite similar to the last study as we also investigate the effect of state variables on the probability of a goal being scored.

4. Strategy in Soccer: An Empirical Study

In this paper, we focus on the dynamic nature of a soccer (football) game and examine the role of asymmetry in team qualities in determining the probability of a goal. Let us first lay out the structure of game play in a football match. A typical football match is a contest between two sets of 11 players who are managed by a professional team manager. The football match consists of two halves of 45 minutes each during which the clock does not stop. Time wasted by players or time taken up by injuries is added on after the end of each half by the fourth official. There is usually a bench constituting of substitute players for each team, out of which 3 players can be brought on by the manager during the course of the game for injury or strategic reasons. The team that scores the most goals in the stipulated 90 minutes (plus added time) wins the match. From this description it is clear that football matches provide us with concrete situations to empirically test game-theoretic predictions in a dynamic setting, given that players operate under a strict set of rules and outcomes for every match can be easily measured.

This setup closely resembles a 2-player contest as mentioned in the summary of literature above. Players spend efforts, which are costly and sunk. Higher efforts increase the probability of winning the contest but the team that puts in the highest total effort is not assured of a win. The probability of winning, in turn, is dependent on the probability of scoring a goal. In case the team does not win, efforts are wasted. The Tullock-type contest would be an ideal setup for this study, but for such an analysis we need sophisticated data which allows us to calculate the effort level of each team. In the absence of such data, we have to carry out the analysis with observables such as number of goals, timing of a goal, state of the game and difference in skill levels. Instead we choose to use the model developed by Moschini (2010) which creates a one-period model, the observable implications of which extend to the multi-period setup. Since this paper adopts the dynamic minute-by-minute approach, we use the Moschini (2010) model and refine it to suit our purposes by using the parametric specification.

Some soccer papers provide us with an appealing structure for a soccer match. The paper by Brocas and Carrilo (2004) was the first one to come up with a theoretical model of the soccer game. Moschini (2010) builds on this and provides a simple game-theoretic representation of the strategic situation in a soccer match. Although these papers provide us with explicit models, they are yet to be tested with real world data. Our motivation for writing this paper stems mainly from this. We use the

models developed in Moschini (2010) and Brocas and Carrillo (2004) to derive comparative statics predictions and test them with real world data.

One should, however, bear in mind the limitations of applying game theoretic models to soccer. While strategies in the form of best response functions and equilibrium action profiles form the basis of game-theoretic analysis, such variables are not observable in real world data. In order to measure the actual degree of attack of any team, we need sophisticated data which tells us the actual location of players at all times during a game which might help us calculate the level of attack and defense by forming an index. However, for our exercise, such data is unavailable, so we approach the problem from observable data such as data on goals scored. This is the reason that we motivate the testable hypotheses via a numerical exercise.

5. A Strategic Model of Soccer

We use the model specification developed by Moschini (2010). Each team chooses an action $a_i \in [0,1]$ where $i \in \{A, B\}$ represent the teams. a_i represents the degree of attacking play. $a_i \rightarrow 1$ indicates towards more attacking play while $a_i \rightarrow 0$ means more defensive play. $p_i(a_i, a_j)$ represents the probability of team i scoring a goal in one-period given that team j does not score in that period. $p_i(a_i, a_j)$, naturally, is a function of the action profile chosen by both players denoted by (a_i, a_j)

The following restrictions on the goal-scoring probability are assumed:

$$\frac{\partial p_i}{\partial a_i} \geq 0, \frac{\partial p_i}{\partial a_j} \geq 0, \frac{\partial^2 p_i(a_i, a_j)}{\partial a_i^2} \leq 0 \text{ and } \frac{\partial^2 p_i(a_i, a_j)}{\partial a_j^2} \geq 0 .$$

The main idea behind the sign of the first two partial derivatives is that teams are more likely to score when they attack. More attack also makes teams vulnerable in defense and hence when the other team attacks, the chances of a particular team scoring at the other end, increases too.

Time of the match is denoted by: $t = \{1, 2, \dots, T\}$ where $T = 90$.

5.1: 1-Period Model

To motivate and introduce the multi-stage model, we start with a one-period model. We assume that the team that wins the one-period contest wins 3 points, the team that loses wins 0 points and both teams get 1 point each in the event of a draw. Assuming a_i and a_j to be the effort levels of the two teams, the payoff to player i in the one-period game is given by:

$$U_i(a_i, a_j; n) = p_i(a_i, a_j) \cdot 3 + p_j(a_j, a_i) \cdot 0 + [1 - p_i(a_i, a_j) - p_j(a_j, a_i)] \cdot 1$$

Which yields

$$U_i(a_i, a_j; n) = 1 + 2 \cdot p_i(a_i, a_j) - p_j(a_j, a_i); i, j \in \{A, B\}, j \neq i$$

Payoffs for teams 1 and 2 are:

$$U_1(a_1, a_2) = 1 + 2 \cdot p_1(a_1, a_2) - p_2(a_2, a_1)$$

$$U_2(a_1, a_2) = 1 + 2 \cdot p_2(a_2, a_1) - p_1(a_1, a_2)$$

We use a fairly general parameterization of the probability function which satisfies the assumptions laid out before and allows for strategic interaction between the action levels of the two participants. Specifically,

$$p_i = \gamma \left[(1 + \beta) \left(a_i - \frac{\lambda}{2} a_i^2 \right) + \frac{\phi}{2} a_j^2 + \eta a_i a_j \right]$$

$$p_j = \gamma \left[(1 - \beta) \left(a_j - \frac{\lambda}{2} a_j^2 \right) + \frac{\phi}{2} a_i^2 + \eta a_i a_j \right]; i, j \in \{1, 2\}, j \neq i$$

where β_i is a measure of strength of a team. The Skill-Difference parameter is defined as $\beta = \beta_i - \beta_j$

This parameter is the source of asymmetry between the teams and the method of deriving this variable for our model is explained below. γ is a scaling parameter that ensures that probabilities lie in the unit interval.

The above parameterization encompasses the one used in Brocas and Carrilo (2004) for $\lambda = 0, \eta = 0$ and $\beta_i = \beta_j$. We choose the five parameters in our parametric model $\lambda, \beta, \gamma, \phi$ and η carefully to deduce Markov Perfect Nash Equilibria of this game. The following restrictions on the parameter values are imposed for regularity conditions:

1. $0 < \beta < 1$ for the probabilities to be increasing in own action
2. $\lambda \leq 1$ for monotonicity for all $a_i \in [0, 1]$
3. $\phi > 0$ and $\eta > 0$ to allow for strategic interaction between the players *i.e.* to ensure that

$$\frac{\partial p_i}{\partial a_j} \geq 0$$

4. Finally, since $p_A + p_B = \gamma[2 - \lambda + \phi + 2\eta]$, we require $\gamma[2 - \lambda + \phi + 2\eta] \leq 1$ for a non-negative draw probability when evaluated at $a_{i2} = a_{j2} = 1$
5. It is noted that the second order condition for maximization is satisfied

$$\text{Since } \frac{\partial^2 U_{i2}}{\partial a_{i2}^2} = -\gamma\lambda(n-1)(1+\beta) - \phi < 0 \text{ for positive } \lambda, \beta, \gamma, \phi.$$

Without loss of generality, we assume that teams are numbered such that the i -th team is stronger *i.e.* $\beta_i > \beta_j$ and $\beta > 0$.

The payoffs of the two players, using the parameterization described above are:

$$U_i = 1 + 2\gamma[(1+\beta)(a_i - \frac{\lambda}{2}a_i^2) + \frac{\phi}{2}a_j^2 + \eta a_i a_j] - \gamma[(1-\beta)(a_j - \frac{\lambda}{2}a_j^2) + \frac{\phi}{2}a_i^2 + \eta a_i a_j]$$

$$U_j = 1 + 2\gamma[(1-\beta)(a_j - \frac{\lambda}{2}a_j^2) + \frac{\phi}{2}a_i^2 + \eta a_i a_j] - \gamma[(1+\beta)(a_i - \frac{\lambda}{2}a_i^2) + \frac{\phi}{2}a_j^2 + \eta a_i a_j]$$

First order conditions:

$$\frac{\partial U_i}{\partial a_i} = 0 \Rightarrow 2.[(1+\beta)(1-\lambda a_i) + \eta a_j] - [\phi a_i + \eta a_j] = 0$$

$$\frac{\partial U_j}{\partial a_j} = 0 \Rightarrow 2.[(1-\beta)(1-\lambda a_j) + \eta a_i] - [\phi a_j + \eta a_i] = 0$$

Which yields:

$$2(1+\beta) - a_i(2\lambda + \phi) + \eta a_j = 0$$

$$2(1-\beta) - a_j(2\lambda + \phi) + \eta a_i = 0$$

The Nash Equilibrium action profile in the static game:

$$a_i^* = \frac{2}{(2\lambda + \phi - \eta)} + \frac{2\beta}{(2\lambda + \phi + \eta)}$$

$$a_j^* = \frac{2}{(2\lambda + \phi - \eta)} - \frac{2\beta}{(2\lambda + \phi + \eta)}$$

Subject to the boundary conditions: $0 \leq a_i^*, a_j^* \leq 1$

We are interested in how the action profile changes when the skill difference between the teams is varied:

$$\frac{\partial a_i^*}{\partial \beta} = \frac{2}{(2\lambda + \phi + \eta)} > 0$$

$$\frac{\partial a_j^*}{\partial \beta} = -\frac{2}{(2\lambda + \phi + \eta)} < 0$$

Thus in the one-period model it is clear that the equilibrium action level of the stronger team increases as the skill difference between the two teams increases. The equilibrium action level of the weaker team decreases as the skill difference between the two teams increases. Thus teams become more attacking as they become stronger and weaker teams become more defensive. These results are clear in the static game. Since we are interested in the dynamic nature of the soccer game, we intend to use the structure of this one-period model to create a multi-period model where the state of the game after the end of the previous period becomes a crucial parameter.

5.2 Multi-stage model ¹:

For each stage we assume the probability parameterization used in Moschini (2010) paper, i.e.

$$p_i = \gamma[(1 + \beta)(a_i - \frac{\lambda}{2}a_i^2) + \frac{\phi}{2}a_j^2 + \eta a_i a_j]$$

$$p_j = \gamma[(1 - \beta)(a_j - \frac{\lambda}{2}a_j^2) + \frac{\phi}{2}a_i^2 + \eta a_i a_j]; i, j \in \{1, 2\}, j \neq i$$

Because each stage can have at most one winner, and a draw is possible in each stage, for each player only one goal can be scored at any one stage. For a player $i \in \{1, 2\}$ there are $2N + 1$ terminal states

¹ This section is adapted from notes provided by Professor Moschini.

$s_i \in \{-N, -(N-1), \dots, -1, 0, 1, \dots, N-1, N\}$. At any stage t there are only $2(t-1)+1$ possible states, depending on how many goals have been scored up to that point. Of course, at any point of the game it must be that $s_i + s_j = 0, i, j \in \{1, 2\}, j \neq i$;

Let:

$U_{i,t}^{(s_i)}$ = Payoff of player i in stage t , with initial state at that stage being s_i

$a_{i,t}^{(s_i)}$ = Action by player i in stage t , with initial state at that stage being s_i

$p_{i,t}^{(s_i)}$ = Scoring probability of player i in stage t , implied by the action profile $(a_{i,t}^{(s_i)}, a_{j,t}^{(s_j)})$

Consider any stage t and state s for player i (state $-s$ for player j). The payoff function of player i is:

$$U_{i,t}^{(s)} = p_{i,t}^{(s)} U_{i,t+1}^{(s+1)*} + p_{j,t}^{(-s)} U_{i,t+1}^{(s-1)*} + (1 - p_{i,t}^{(s)} - p_{j,t}^{(-s)}) U_{i,t+1}^{(s)*}$$

$$\text{or, } U_{i,t}^{(s)} = U_{i,t+1}^{(s)*} + p_{i,t}^{(s)} (U_{i,t+1}^{(s+1)*} - U_{i,t+1}^{(s)*}) - p_{j,t}^{(-s)} (U_{i,t+1}^{(s)*} - U_{i,t+1}^{(s-1)*})$$

Hence:

$$U_{i,t}^{(s)*} = U_{i,t+1}^{(s)*} + p_{i,t}^{(s)} \Delta_{i,t,(s)}^+ - p_{j,t}^{(-s)} \Delta_{i,t,(s)}^-$$

Where

$$\Delta_{i,t,(s)}^+ = U_{i,t+1}^{(s+1)*} - U_{i,t+1}^{(s)*} \text{ ie. the "up" change in equilibrium payoffs}$$

$$\Delta_{i,t,(s)}^- = U_{i,t+1}^{(s)*} - U_{i,t+1}^{(s-1)*} \text{ ie. the "down" change in equilibrium payoffs}$$

Simplifying notation, we can alternatively write payoff at time t as:

$$V_{i,t} = V_{i,t+1} + p_{i,t} \Delta_{i,t}^+ - p_{j,t} \Delta_{i,t}^-$$

$$V_{j,t} = V_{j,t+1} + p_{j,t} \Delta_{j,t}^+ - p_{i,t} \Delta_{j,t}^-$$

Given the parameterization of probabilities, Nash equilibrium conditions are:

$$\frac{\partial V_i}{\partial a_i} = \Delta_{i,t}^+ \cdot [(1 + \beta)(1 - \lambda a_i) + \eta a_j] - \Delta_{i,t}^- \cdot (\phi a_i + \eta a_j) = 0$$

$$\frac{\partial V_j}{\partial a_j} = \Delta_{j,t}^+ \cdot [(1 - \beta)(1 - \lambda a_j) + \eta a_i] - \Delta_{j,t}^- \cdot (\phi a_j + \eta a_i) = 0$$

Which yields:

$$\Delta_{i,t}^+ \cdot (1 + \beta) - a_i (\lambda (1 + \beta) \Delta_{i,t}^+ + \phi \Delta_{i,t}^-) + a_j \eta (\Delta_{i,t}^+ - \Delta_{i,t}^-) = 0$$

$$\Delta_{j,t}^+ \cdot (1 - \beta) - a_j (\lambda (1 - \beta) \Delta_{j,t}^+ + \phi \Delta_{j,t}^-) + a_i \eta (\Delta_{j,t}^+ - \Delta_{j,t}^-) = 0$$

Or:

$$\begin{pmatrix} A_i & -B_i \\ -B_j & A_j \end{pmatrix} \begin{pmatrix} a_i \\ a_j \end{pmatrix} = \begin{pmatrix} C_i \\ C_j \end{pmatrix}$$

Where:

$$A_i = \lambda (1 + \beta) \Delta_{i,t}^+ + \phi \Delta_{i,t}^-$$

$$A_j = \lambda (1 - \beta) \Delta_{j,t}^+ + \phi \Delta_{j,t}^-$$

$$B_i = \eta (\Delta_{i,t}^+ - \Delta_{i,t}^-)$$

$$B_j = \eta (\Delta_{j,t}^+ - \Delta_{j,t}^-)$$

$$C_i = \Delta_{i,t}^+ (1 + \beta)$$

$$C_j = \Delta_{j,t}^+ (1 - \beta)$$

Letting $\det = A_i A_j - B_i B_j$ and solving by Cramer's rule, we derive the Markov Perfect Nash

Equilibrium (MPNE) action profile in the multi-stage game as:

$$a_i^* = \frac{A_j C_i + B_i C_j}{\det} \quad \text{and} \quad a_j^* = \frac{A_i C_j + B_j C_i}{\det} \quad \text{subject to the boundary conditions } 0 \leq a_i^*, a_j^* \leq 1$$

The algorithm used for simulations to derive MPNE action profile described in the section below is :

- i. Solve for (a_i^*, a_j^*) using formulae for interior solution above. If $0 \leq a_i^*, a_j^* \leq 1$ then end.
- ii. If both violate boundary conditions, set both values at the corresponding boundary
- iii. If one of the (a_i^*, a_j^*) violates boundary, set it to boundary value and find the other value

from the Best Response functions: $a_i^* = \frac{C_i + a_j B_i}{A_i}$ and $a_j^* = \frac{C_j + a_i B_j}{A_j}$, where the other action $\in \{0,1\}$.

5.3 Simulations:

Because general analytical results do not appear possible in this setup, the conjectures that are tested empirically are motivated with the help of simulations. To start, we selected a set of parameters that satisfy the various regularity conditions. The initial parameter values are:

$\gamma = 0.25, \beta = 0, \lambda = 1, \phi = 1$ and $\eta = 0.4$. It may be noted that $\beta = 0$ denotes the symmetric case, where the skill difference between teams is zero. This parameter vector satisfies the regularity conditions mentioned above and are used to calculate the Markov-Perfect Nash Equilibria in a 6-period game.

The two-period game described above gives us important results that can be used in the 6-period game. To begin with, conditional on the state of the game being tied, the last period in an N -period game is the same, regardless of the value of N . Therefore, if teams are tied at the beginning of the sixth period, equilibrium action profiles in that period are exactly the same as that in the one-period model, because in the one period model teams always start with a tie. Also, if the i -th team is trailing at the beginning of the sixth period, then computation of MPNE action profile is trivial, given by $a_{i6}^* |_{s_i=1} = 1$ and $a_{j6}^* |_{s_i=1} = 0$, i.e., the trailing team puts in maximum attack and the leading team goes into full defense in the last period.

5.4 Simulation results and conjectures:

I calculate Markov-Perfect-Nash Equilibrium (MPNE) action profiles for each team in each period of a six-period game with the vector of parameters above. Simulation results are reported in tables 1 to 6.

I consider three cases –

- a) When the state of the game is a tie
- b) When the stronger team is trailing by a goal.
- c) When the stronger team is leading by a goal

Given that the i -th team is stronger, these three scenarios represent the most general situations in soccer matches.

Since I have already derived expressions for the MPNE action profile in each period of the six-stage game, I use them along with the set of parameters chosen to calculate the actual equilibrium action profiles. I first carry out the exercise in the symmetric case, with $\beta = 0$. Next, I use the MPNE action profile values to calculate $p_i(a_{i,k}^*, a_{j,k}^*)$ and $p_j(a_{j,k}^*, a_{i,k}^*)$, $k = 1, 2, \dots, 6$. After calculating these probabilities I am able to calculate the probability of observing a goal in periods 1 to 6 by calculating $p_i(a_{i,k}^*, a_{j,k}^*) + p_j(a_{j,k}^*, a_{i,k}^*)$.

Tables 1 and 2 report values conditional on the state of the game being a tie. Tables 3 & 4 do the same conditional on the state of the game being that the stronger team is trailing by one goal and the j -th team leading by a goal. Tables 5 & 6 do the same conditional on the state of the game being that the stronger team is leading by a goal whereas the weaker team is trailing by a goal. Tables 1 to 6 are used to motivate conjectures the following conjectures. All conjectures assume that other explanatory variables are held constant.

Conjecture 1: Conditional on the state of the game being a tie, the probability of observing a goal increases as the end of the game approaches

Simulation results for the probability of observing a goal given that the state of the game is a tie are reported in table 2. It is clear from table 2 that no matter what the absolute value of the skill difference between the teams is, the probability of observing a goal goes up over time as the game progresses, given that it is tied.

Conjecture 2: Probability of observing a goal is higher if game is tied versus if it is not tied.

For this conjecture, simulation results reported in tables 2, 4 and 6 are compared. Tables 4 and 6 report total probability of observing a goal given that the game is not tied for two symmetrically opposite situations. Simple comparison of numbers between these tables tell us that numbers reported in table 2 are larger than the other two. This indicates that the probability of observing a goal is higher if the game of soccer is tied versus if it is not.

Conjecture 3: Probability of observing a goal is higher if the weaker team leads compared to the situation where the stronger team leads

Tables 4 and 6 are relevant for this conjecture. Comparison between these two tables shows that numbers in table 4 are larger than those in table 6 which motivates this conjecture.²

Conjecture 4: Conditional on the state of the game being that the game is tied, the probability of observing a goal is insensitive to the skill difference between teams

Table 2 is relevant for conjecture 4. Rows in this table are in increasing order of skill difference from top to bottom. For all columns p_1 to p_6 , increasing values of skill difference between teams (due to increases in the absolute value of skill difference between teams, $\beta = \beta_i - \beta_j$) do not lead to significant changes in the probability of observing a goal in any period. This motivates conjecture 4.

Conjecture 5: Final period effect - Probability of observing a goal is significantly higher in the final period of the game compared to the second last period, regardless of the state of the game

Tables 2, 4 and 6 are relevant for this conjecture. For all three cases, there is a jump in probabilities in the last period, compared to the second last period.

These conjectures are tested using a large dataset which was created using information on matches played in the Italian Serie A and the Spanish La Liga which are top level professional leagues in two countries that are giants of international football – Italy and Spain. We are interested in conjectures that deal with the probability of observing goals in a game of soccer and not actions, because actions are not observable while goals are.

6. A Strategic Model of Soccer: Empirical Results

6.1 Data and descriptive statistics:

In this paper I look at the soccer game in a unique way by visualizing each minute (1 - 90) as a stage game. I use a minute-by-minute approach where in each minute at most 1 goal can be scored by either team. In a typical soccer match, players choose strategies every minute after taking into careful consideration the state of the match. The 'state' may be measured by how many goals a team is ahead or behind, or by how much time is left to play. In order to better understand what constitutes 'optimal

² This excludes the symmetric case where the teams have no skill difference, in which case there is no strong or weak team.

strategies' in a dynamic contest such as a soccer match, I think that the soccer match can be envisaged as a dynamic game where teams implicitly change strategies every minute through their style of play, depending on the state of the match. Since style of play and level of attack cannot be measured, we focus on measurable outcomes such as goals and propose a unique method of analysis where equilibrium strategies result in outcomes such as goals. The level of attack and defense chosen by teams is what leads to goals, and conjectures that are derived using a simulation technique which takes into account the state the skill difference between teams it tested with data on the only observable outcome – goals.

The reaction of teams to incentives such as rewards for a win or their choice of optimal strategies, given the state of the game, can only be analyzed in a dynamic setting. Previous work in this area has allowed for a change in strategy only twice per match - at the beginning of the game and at the beginning of the second half. In truth, managers make subtle changes to their strategies every minute in a match. These strategies are communicated to players during training drills and professional players carry them out during live play. If the manager changes strategies during a game, then the manager tries to communicate it to players from the sidelines. If the manager feels that the team is not playing according to the strategy decided by the manager and the coaching staff, then either players get yelled at or they are substituted from the game and eventually face disciplinary action, which can lead to termination of player contracts or the player being left out of the team in future matches. The level of attack and defense chosen by teams are also dependent on the state of the match and the probability of a goal being scored, in turn, is determined by the state of the game and the amount of time left in the match. This paper is the first of its kind, one that incorporates a minute-by-minute analysis to soccer.

The role of fatigue, asymmetry in quality of teams and the exact state of the match on the probability of a goal being scored are also analyzed in this paper. I do not try to measure 'strategy'. The aim is to test 6 conjectures derived using data on observable outcomes – goals.

The main source of data is the RSSSF website at <http://rssf.com>. RSSSF stands for Rec.Sport.Soccer Statistics Foundation. It is a 20-year-old online archive that was formed in 1994. The website contains league tables and other statistical information on football. To provide some background on our source - rssf.com initially started as a Northern European organization but has since spread to include statistical information from football matches all over the world. I use this website to compile minute-by-minute data from yearly information after extensive data manipulation.

A snapshot of the data for Italy for the 2009/10 season on the RSSSF website is provided in Figure 1. As is clear from this figure, data on this website appears round wise for each match. For any line of data, the first word refers to the name of the home-team, followed by the final score of the match and then the away team name. The next line provides information on goals and goal scorers. This line mentions the name of the goal scorer followed by the exact minute when the goal was scored. Commas separate multiple goal scorers from the same team whereas semicolons separate goal scorers from different teams. This was the raw data that I used. I used extensive data manipulation techniques using regular expressions in the statistical software R to transform this raw data into the final primitive data, a snapshot of which may be seen in Table 19.³ I use data from 7 seasons from the Italian Serie A (2005 - 2011) and 5 seasons from the Spanish La Liga (2003 to 2007). Both leagues have 20 participating teams, playing against each other in a league format with the winner getting 3 points and both teams getting 1 point in case the game finishes a tie. Any pair of two teams meet twice during a regular season, playing two matches - one at home and one away, for a total of 380 matches per league per season. I analyze a total of 2,680 games from Italy. During this period a total of 6,785 goals were scored at an average of 2.55 goals per match. In Spain the average was 2.59 over 1895 matches in 5 seasons for a total of 4,914 goals. The following tables A and B provides summary statistics for some key variables for Italy and Spain datasets.

Each soccer season is divided into 38 Match Days with 10 games being held on each Match Day which involves all 20 teams. Since we conceive a soccer match as a series of 90 one-minute games, each match is analyzed per minute. This gives rise to $380 * 90 = 34,200$ rows of data for each season. Figures 2 and 3 are histograms of average goals scored by minute in Italy and Spain. There are two peaks seen in this graph at minutes 45 and 90 which illustrate the nature of the data - usually injury time is added at the end of both halves of soccer and these minutes typically represent a period of time longer than one actual minute in our data. Spain, interestingly, does not have a peak in the 90th minute.

Descriptive statistics reported below in tables A and B show that the stronger team wins most games both in Italy and Spain. However, the weaker team beats the stronger team in 24.7% of matches in Italy and 23.6% of matches in Spain. Therefore, the skill difference between teams is not the only factor that determines the result of soccer matches in Spain and Italy. Also, a significant percentage of games end in ties, therefore conjectures derived with the state of the game being a tie are important.

³ The definition of variables used in Table 7 are provided below.

Table A-Summary statistics for Italy (2005 to 2011)

	No. Of Games	% of
Total Games	2680	
Stronger team wins	1227	45.7%
Weaker team wins	664	24.7%
Games ending in ties	789	29.4%
Games with a goal in the 45th minute	199	7.4%
Games with a goal in the 90th minute	414	15.4%

Data source: rsssf.com

Table B-Summary statistics for Spain (2003 to 2007)

	No. of games	% of games
Total games	1895	
Stronger team wins	762	40.2%
Weaker team wins	447	23.6%
Games ending in ties	686	36.2%
Games with a goal in the 45th minute	29	1.5%
Games with a goal in the 90th minute	65	3.4%

Data source: rsssf.com

It may be useful to mention here that data on the minute on which a goal is scored in professional soccer matches was very difficult to find online. Most of the data sources that are available mention the overall score of the match and provide information on goal scorers. However, data on the minute in which goals were scored was hard to find. This could be because such information is stored by data-mining authorities and are not freely available. For example, the website <http://www.espn.co.uk/> keeps information on match results for the English Premier League. Such websites for other leagues can be found too, but they do not contain information on the exact minute of goals scored. Since it is critical for this paper to know which minute the goal was scored, to predict the probability of a goal being scored in a particular minute, the <http://rsssf.com> website data shown in figure 1 is most useful to us despite the fact that compilation of data was an arduous task. I also use data on seasonal performances taken from the final league tables, also available on this website, to create a measure of asymmetry between teams by a method which is described in detail below.

6.2 Validation:

To validate the accuracy of the data, I recreated the final league tables for every year using the minute-by-minute dataset. Since 3 points are awarded for a win, 1 for a draw and none for a loss I was able to exactly replicate the final league tables by using home team and away team IDs and the final state of the match.

In order to derive β from the data, I create a unique skill variable and the method used to derive it is explained in the section below.

7. Skill Variable ⁴

As mentioned in the survey of literature above, there is vast literature on contests between asymmetric players. This paper deals with soccer, where teams are of different qualities.⁵ In order to capture the degree of asymmetry between teams, we construct a *Skill* variable that captures the strength of the team at a particular point in time. This section describes how the skill variable is calculated for each team for every match. The strategy that we use is to approximate the unobservable strength of a team by their observed performance in the previous 12 months. To do so, let $r_{i,m}^y \in \{0,1,3\}$ denote the number of points earned by team i in match day m in season y (recall that teams earn 3 points for a win, one point for a tie, and zero points for a loss). Next, we can represent the total points earned by a team in a given

season y , up to and including match day $m = \{1, 2, \dots, M\}$ as: $P_{i,m}^y = \sum_{d=1}^m r_{i,d}^y$.

7.1 Initial Team Strength: based on the foregoing, a team's strength at the beginning of a new season (i.e.,

before the first match day), is represented as $s_{i,1}^y = \frac{P_{i,M}^{y-1}}{3M}$. Note that this is equal to the points that a team

earned in the previous season as a fraction of the total possible points (which is $3M$, where M is the total number of match days in a season). For newly promoted teams, their starting strength is set equal to that of the last team in the standing not to be relegated.

7.2 Updating Teams' Strength Measures: The initial measure of a team's strength is updated throughout the season, after each match day, based on the team's performance. The presumption is that

⁴ This section has been adapted from notes provided by Prof Moschini.

⁵ Quality difference may arise due to financial power or history, reputation, popularity of the club or the quality of its academy etc.

this will reflect new information which may be related to player injuries, arrival of new players after transfer windows, coach changes etc. The updating formula is as follows:

$$s_{i,m}^y = \frac{m-1}{M} \left(\frac{P_{i,m-1}^y}{3(m-1)} \right) + \left(1 - \frac{m-1}{M} \right) \cdot \frac{P_{i,M}^{y-1}}{3M}; m = 2, 3, \dots, M$$

With the above formulation, after the last match of the season, strength is defined only on the team's performance during that season:

$$s_{i,M+1}^y = \frac{1}{3M} \sum_{d=1}^m r_{i,d}^y = \frac{P_{i,M}^y}{3M}$$

This defines the team's initial strength in the next season i.e. $s_{i,1}^{y+1} = s_{i,M+1}^y$. We use this method to derive the *Skill* variable for each team for every match day of the season and use the value of their absolute difference as a measure of asymmetry in skill levels of the two competing teams.

Next, we describe the econometric model that we used for this paper.

Summary statistics for the skill variable are reported in tables 19 and 20. Mean of skill variable is 0.4 for both Spain and Italy although the range of values is large. The absolute value of the skill difference between teams is relevant for this paper. These tables show that there is variation in skill difference between teams with the maximum value reaching 0.48 for Italy and 0.39 for Spain. This shows the importance of using a skill variable which correctly represents difference in quality of teams.

8. Econometric Model and Results

We report estimation results for two econometric models are used to test conjectures 1 to 5.

I use the following baseline Probit specification:

$$\begin{aligned} Goal_{ijdtm} = & \alpha + \beta_0 SS_{ijdtm} + \beta_1 SW_{ijdtm} + \beta_2 T * SZ_{ijdtm} + \beta_3 T * SW_{ijdtm} + \beta_4 T * SS_{ijdtm} \\ & + \beta_5 SkillDiff * SZ_{ijdtm} + \beta_6 SkillDiff * SW_{ijdtm} + \beta_7 SkillDiff * SS_{ijdtm} + \beta_8 Dum_{-45}_{ijdtm} \\ & + \beta_9 Dum_{-90}_{ijdtm} + \beta_{10} DES_{ijdtm} + \beta_{11} DF_{ijdtm} + \varepsilon_{ijdtm} \end{aligned}$$

where:

i, j = Team IDs

d = Match Day (=1,2,...,38)

t = year

T = minutes

$Goal$ = Dummy for a goal being scored in that minute (value=1 if goal scored in that minute, 0 otherwise)

SZ = Dummy for “state” being a tie (value=1 if match is tied, 0 otherwise)- This is used as the baseline state and is excluded.

SW = Dummy for “state” being weaker team is leading (value=1 if weaker team leads, 0 otherwise)

SS = Dummy for “state” being leading team is stronger or of equal skill as the other team (value=1 if stronger team (or teams of equal skill) is leading, 0 otherwise)

$SkillDiff$ = Absolute value of Skill Difference between the two teams

$Dum45$ = Dummy for last minute of First half (=1 if $T=45$, 0 otherwise)

$Dum90$ = Dummy for last minute of Second half (=1 if $T=90$, 0 otherwise)

DES = Dummy for “end of season” effect (=1 if $d \geq 35$)

DF = Dummy for “fatigue” effect (=1 if $T \geq 61$)

ε_{ijdtm} is the error term.

Estimation results are reported in tables 7 and 8. For both Italy and Spain, we report estimation results for two models. Pairwise likelihood ratio tests were carried out for model selection for both Spain and Italy and these two models were selected.

8.1 Italy results:

Estimation results are reported in table 7. Model 1 is the general model which uses the dummy for fatigue as well as yearly dummies as explanatory variables. Model 2 drops the fatigue variable and yearly dummies.

For both models 1 and 2, where the minute variable is interacted with the state of the game, coefficients for all three interaction terms are positive and significant. These are strong results because they prove that the state of the game matters. For both models, the coefficient for the *SkillDiff * SZ* term is positive and significant. The *Dum45* and *Dum90* variables also turn out to be highly significant in both models. These minutes are usually longer because of the nature of our data. Also as figure 2 shows, there are peaks in the average number of goals scored in these minutes, so this result is not surprising. The variable for end of season, *DES* also turns out to be highly significant in both models. This suggests that as the end of season approaches, i.e. for the last four matches in a season, the probability of observing a goal is significantly higher than matches played towards the beginning of the season. This might be due to the nature of design of the Italian Serie A (and other professional leagues across the world), where some teams are engaged in a race for the title at the top of the table, while teams at the bottom are engaged in a relegation battle towards the end of a season, both of which require more risk-taking than usual, which probably leads to higher probability of observing goals in these matches. The fatigue variable, *DF* is not significant in model 1.

8.2 Spain results:

Estimation results are reported in table 8.

Results for Spain are mostly similar to Italy although there are a few differences. The minute variable, when interacted with the state of the game does not lead us to conclude that that the state of the game has significant effect on the probability of observing a goal. The *SkillDiff * SZ* variable is, however, still significant at the 5% level of significance. *Dum45* and *DES* are both significant and both have significant positive effects on the probability of observing a goal. Yearly dummies are significant for Spain. These Estimations results help in confirming/rejecting conjectures 1-5 that were derived using simulation results.

Conjecture 1:

Italy: The coefficient of interest here is T^*SZ . This coefficient is highly significant across all the models. Also, the sign of the coefficient is positive, which suggests that it provides support to conjecture 1.

Spain: For Spain, the coefficient for T^*SZ is not significant and hence we are unable to comment on the validity of conjecture 1 for Spain.

Conjecture 2:

The coefficients of interest here are SS and SW , which denote the state of the game being that the stronger and weaker teams are leading, respectively. If both of them are negative and significant, then the hypothesis that the probability of observing a goal is higher if game is tied versus if it is not tied is supported, since we treat the game being tied as the baseline case.

For Italy there is partial support to this conjecture since the coefficient for SS is negative and significant in model 2. Also, the conjecture is not supported for the case where the weaker team leads in Italy. For Spain, again, coefficients are not significant.

Conjecture 3:

This conjecture is supported for Italy since the coefficient for SW is positive and significant whereas the same for SS is negative and significant in Model 2. It is not supported for Spain.

Conjecture 4:

The coefficient of interest here is $SkillDiff * SZ$. For both Italy and Spain this conjecture does not hold.

Conjecture 5:

There is significant final period effect in the sense that the 90th minute dummy appears positive and significant for Italy in both models. However, we are interested in testing whether it is significantly different from the second last period. In order to test that out, we test if standard errors of the 45th and 90th minute dummies differ significantly or not. For both Italy and Spain, test results suggest that the two dummies are different, providing support to conjecture 5.

Thus, we are able to confirm some predictions of simulations and are able to validate some conjectures, which can be summarized as:

- a) Probability of observing a goal goes up as the game of soccer progresses from the first minute towards the 90th minute, given that the game is tied
- b) Probability of observing a goal in soccer is higher if the weaker team leads the game compared to the situation where the stronger team leads
- c) There are significant final period effects in soccer

9. Conclusions

In this paper I use a unique minute-by-minute approach to model a soccer game. An explicit game-theoretic model is used to derive conjectures which are then test empirically using a unique dataset which spans two countries and a number of years. This paper is the first one to look at how the probability of a goal being observed in a game changes as the game progresses.

I consider the ‘state’ of the game to be an important determinant of the probability of observing a goal. The theoretical model helps in identifying Markov Perfect Nash Equilibrium(MPNE) action profile in a multi-stage game, which are used for simulations. The simulation results help in motivating six conjectures. The results of the empirical analysis show that conditional on the state of the game being a tie, the probability of observing a goal goes up as the match progresses. We are also able to show that if the game occurs towards the end of the soccer season, when teams are either fighting for the title or fighting for survival, then matches may have more goals than usual. We also derive a unique skill variable which is used as an instrument for variation in skill levels. Further work in this area may look at the effect of home and away teams on the probability of observing goals.

This paper contributes to the literature on dynamic asymmetric contests. This paper also provides empirical support to the notion of Nash equilibrium. These results may help soccer managers strategize in a more efficient manner, given their knowledge of the state of the game and the quality of their opponent.

TABLE 1

Simulation results – MPNE action profiles,
conditional on the **state of the game being a tie:**

γ	β	λ	Φ	η	ai1	ai2	ai3	ai4	ai5	ai6
0.3	0	1	1	0.4	0.5244	0.5303	0.5408	0.5616	0.6052	0.7692
0.3	0.1	1	1	0.4	0.5301	0.5376	0.5502	0.5734	0.6193	0.7818
0.3	0.2	1	1	0.4	0.5344	0.5435	0.5583	0.5841	0.6322	0.7926
0.3	0.3	1	1	0.4	0.5377	0.5486	0.5658	0.594	0.6443	0.8019
0.3	0.4	1	1	0.4	0.5402	0.5532	0.5728	0.6035	0.6557	0.8098
0.3	0.5	1	1	0.4	0.5424	0.5577	0.5798	0.6129	0.6669	0.8163
0.3	0.6	1	1	0.4	0.5449	0.5627	0.5873	0.6227	0.678	0.8216
0.3	0.7	1	1	0.4	0.5485	0.5689	0.596	0.6331	0.6892	0.8256
0.3	0.8	1	1	0.4	0.5544	0.5772	0.6062	0.6448	0.7008	0.828
0.3	0.9	1	1	0.4	0.5642	0.5887	0.6191	0.6581	0.713	0.8286
γ	β	λ	Φ	η	aj1	aj2	aj3	aj4	aj5	aj6
0.3	0	1	1	0.4	0.5244	0.5303	0.5408	0.5616	0.6052	0.7692
0.3	0.1	1	1	0.4	0.517	0.5214	0.5299	0.5483	0.5895	0.7545
0.3	0.2	1	1	0.4	0.5075	0.5104	0.5169	0.533	0.5718	0.7373
0.3	0.3	1	1	0.4	0.4953	0.4966	0.5012	0.515	0.5514	0.717
0.3	0.4	1	1	0.4	0.4797	0.4793	0.4818	0.4935	0.5276	0.6927
0.3	0.5	1	1	0.4	0.4594	0.4572	0.4579	0.4673	0.4991	0.6633
0.3	0.6	1	1	0.4	0.4326	0.4287	0.4276	0.4347	0.4645	0.627
0.3	0.7	1	1	0.4	0.3966	0.3909	0.3883	0.3933	0.4214	0.5814
0.3	0.8	1	1	0.4	0.3469	0.34	0.3358	0.3392	0.366	0.5223
0.3	0.9	1	1	0.4	0.2764	0.2687	0.2635	0.266	0.2924	0.4429

TABLE 2

Simulation results –Probability of observing a goal in each period of a six-period game, conditional on the **state of the game being a tie**:

γ	β	λ	Φ	η	p1	p2	p3	p4	p5	p6
0.3	0	1	1	0	0.317	0.321	0.329	0.344	0.376	0.503
0.3	0.1	1	1	0	0.317	0.321	0.329	0.344	0.376	0.502
0.3	0.2	1	1	0	0.315	0.32	0.328	0.343	0.375	0.5
0.3	0.3	1	1	0	0.313	0.318	0.326	0.341	0.373	0.496
0.3	0.4	1	1	0	0.31	0.315	0.323	0.339	0.37	0.491
0.3	0.5	1	1	0	0.306	0.311	0.32	0.336	0.367	0.483
0.3	0.6	1	1	0	0.3	0.306	0.316	0.332	0.362	0.473
0.3	0.7	1	1	0	0.294	0.301	0.311	0.327	0.357	0.46
0.3	0.8	1	1	0	0.287	0.294	0.305	0.321	0.349	0.444
0.3	0.9	1	1	0	0.279	0.287	0.298	0.314	0.34	0.423

TABLE 3

Simulation results – MPNE action profiles, conditional on the state of the game being that the stronger team is trailing (weaker team leading)

γ	β	λ	Φ	η	ai2	ai3	ai4	ai5	ai6
0.25	0	1	1	0.4	0.7479	0.7895	0.832	0.9001	1
0.25	0.1	1	1	0.4	0.7485	0.7901	0.8337	0.9015	1
0.25	0.2	1	1	0.4	0.7492	0.7904	0.8346	0.9027	1
0.25	0.3	1	1	0.4	0.7499	0.7906	0.8349	0.9038	1
0.25	0.4	1	1	0.4	0.7506	0.7906	0.8347	0.9048	1
0.25	0.5	1	1	0.4	0.7514	0.7907	0.8341	0.9057	1
0.25	0.6	1	1	0.4	0.7523	0.7905	0.8343	0.9063	1
0.25	0.7	1	1	0.4	0.7532	0.79	0.8364	0.9068	1
0.25	0.8	1	1	0.4	0.7542	0.7893	0.8383	0.9069	1
0.25	0.9	1	1	0.4	0.7549	0.7917	0.84	0.9067	1
γ	β	λ	Φ	η	aj2	aj3	aj4	aj5	aj6
0.25	0	1	1	0.4	0.2327	0.1582	0.0503	0	0
0.25	0.1	1	1	0.4	0.2297	0.1571	0.0511	0	0
0.25	0.2	1	1	0.4	0.2224	0.1514	0.0479	0	0
0.25	0.3	1	1	0.4	0.2107	0.1413	0.0404	0	0
0.25	0.4	1	1	0.4	0.1942	0.1264	0.0288	0	0
0.25	0.5	1	1	0.4	0.1723	0.1065	0.0127	0	0
0.25	0.6	1	1	0.4	0.1444	0.0809	0	0	0
0.25	0.7	1	1	0.4	0.1092	0.0489	0	0	0
0.25	0.8	1	1	0.4	0.0653	0.0095	0	0	0
0.25	0.9	1	1	0.4	0.0105	0	0	0	0

i-th team is stronger

TABLE 4

Simulation results: Probability of observing a goal in each period of a six-period game, conditional on the state of the game being that the **stronger team is trailing (weaker team leading)**

γ	β	λ	Φ	η	p2	p3	p4	p5	p6
0.25	0	1	1	0.4	0.2799	0.2619	0.2289	0.225	0.25
0.25	0.1	1	1	0.4	0.2856	0.2699	0.2406	0.2377	0.2625
0.25	0.2	1	1	0.4	0.2898	0.2763	0.2506	0.2504	0.275
0.25	0.3	1	1	0.4	0.2927	0.2813	0.2591	0.2631	0.2875
0.25	0.4	1	1	0.4	0.2947	0.2852	0.2665	0.2758	0.3
0.25	0.5	1	1	0.4	0.2958	0.2883	0.273	0.2884	0.3125
0.25	0.6	1	1	0.4	0.2962	0.2907	0.2815	0.3009	0.325
0.25	0.7	1	1	0.4	0.2962	0.2927	0.2943	0.3134	0.3375
0.25	0.8	1	1	0.4	0.296	0.2949	0.307	0.3259	0.35
0.25	0.9	1	1	0.4	0.2963	0.3055	0.3196	0.3382	0.3625

TABLE 5

Simulation results – MPNE action profiles, conditional on the state of the game being that **the stronger team is leading** (weaker team trailing)

γ	β	λ	Φ	η	ai2	ai3	ai4	ai5	ai6
0.25	0	1	1	0.4	0.2327	0.1582	0.0503	0	0
0.25	0.1	1	1	0.4	0.2311	0.1546	0.0452	0	0
0.25	0.2	1	1	0.4	0.2246	0.146	0.0358	0	0
0.25	0.3	1	1	0.4	0.2125	0.1318	0.0219	0	0
0.25	0.4	1	1	0.4	0.1936	0.1111	0.0031	0	0
0.25	0.5	1	1	0.4	0.1662	0.0829	0	0	0
0.25	0.6	1	1	0.4	0.128	0.0466	0	0	0
0.25	0.7	1	1	0.4	0.0763	0.0022	0	0	0
0.25	0.8	1	1	0.4	0.0111	0	0	0	0
0.25	0.9	1	1	0.4	0	0	0	0	0
γ	β	λ	Φ	η	aj2	aj3	aj4	aj5	aj6
0.25	0	1	1	0.4	0.7479	0.7895	0.832	0.9001	1
0.25	0.1	1	1	0.4	0.7472	0.7886	0.8292	0.8987	1
0.25	0.2	1	1	0.4	0.7464	0.7871	0.8248	0.8971	1
0.25	0.3	1	1	0.4	0.7453	0.7844	0.8181	0.8955	1
0.25	0.4	1	1	0.4	0.7435	0.7797	0.8076	0.8938	1
0.25	0.5	1	1	0.4	0.7395	0.7705	0.8025	0.892	1
0.25	0.6	1	1	0.4	0.7307	0.7535	0.7988	0.8902	1
0.25	0.7	1	1	0.4	0.7106	0.7203	0.7949	0.8883	1
0.25	0.8	1	1	0.4	0.6579	0.712	0.7911	0.8864	1
0.25	0.9	1	1	0.4	0.6359	0.7056	0.7874	0.8846	1

i-th team is stronger

TABLE 6

Simulation results: Probability of observing a goal in each period of a six-period game, conditional on the state of the game being that the **stronger team is leading (weaker team trailing)**

γ	β	λ	Φ	η	p2	p3	p4	p5	p6
0.25	0	1	1	0.4	0.2799	0.2619	0.2289	0.225	0.25
0.25	0.1	1	1	0.4	0.2725	0.2518	0.2151	0.2123	0.2375
0.25	0.2	1	1	0.4	0.2629	0.2392	0.1986	0.1995	0.225
0.25	0.3	1	1	0.4	0.2503	0.2232	0.1789	0.1868	0.2125
0.25	0.4	1	1	0.4	0.2338	0.2029	0.1553	0.174	0.2
0.25	0.5	1	1	0.4	0.2118	0.1769	0.1406	0.1612	0.1875
0.25	0.6	1	1	0.4	0.1818	0.1434	0.1277	0.1485	0.175
0.25	0.7	1	1	0.4	0.1402	0.1007	0.1149	0.1357	0.1625
0.25	0.8	1	1	0.4	0.0826	0.0863	0.1021	0.1229	0.15
0.25	0.9	1	1	0.4	0.0614	0.0737	0.0894	0.1101	0.1375

TABLE 7

Probit estimates of the determinants of the probability of observing a goal.

Dependent variable =1 if either team scores

(Italy 2005 to 2011)

	Model 1		Model 2	
	<i>Estimate</i>	<i>P_Value</i>	<i>Estimate</i>	<i>P_Value</i>
SS	-0.0267	0.432	-0.0284	0.033
SW	0.0271	0.494	0.0260	0.039
T*SZ	0.0009	0.005	0.0009	0.000
T*SW	0.0011	0.08	0.0011	0.001
T*SS	0.0021	0	0.0021	0.000
SkillDiff*SZ	0.2258	0.022	0.2186	0.098
SkillDiff*SW	0.2541	0.155	0.2535	0.178
SkillDiff*SS	-0.0119	0.919	-0.0120	0.117
Dum_45	0.4996	0	0.4991	0.037
Dum_90	0.8591	0	0.8596	0.031
DES	0.0637	0	0.0634	0.017
DF	0.0014	0.935		
Dum_2005	0.0026	0.898		
Dum_2006	0.0212	0.286		
Dum_2007	0.0055	0.784		
Dum_2008	0.0031	0.877		
Dum_2009	0.0109	0.584		
Dum_2010	0.0197	0.319		
Intercept	-2.0194	0	-2.0095	0.013
Log-likelihood	-30341.7		-30342.8	
Number of Obs.	239,164		239,164	
Pseudo R2	0.0168		0.0167	

****p-value* < 0.001, ** *p-value* < 0.01, * *p-value* < 0.05

The Y variable was 1 for 6,785 cases. Data source: rssf.com

TABLE 8

Probit estimates of the determinants of the probability of observing a goal.

Dependent variable =1 if either team scores

(Spain 2003 to 2007)

	Model 1		Model 2	
	<i>Estimate</i>	<i>P_Value</i>	<i>Estimate</i>	<i>P_Value</i>
SS	-0.0069	0.855	0.0307	0.393
SW	-0.0562	0.192	-0.0151	0.716
T*SZ	-0.0002	0.511	-0.0003	0.283
T*SW	0.0010	0.166	0.0003	0.685
T*SS	0.0005	0.422	-0.0002	0.739
SkillDiff*SZ	0.2379	0.062	0.2749	0.028
SkillDiff*SW	0.1709	0.44	0.1709	0.440
SkillDiff*SS	0.0941	0.569	0.0941	0.567
Dum_45	0.2973	0	0.3006	0.000
Dum_90	0.0837	0.138	0.0850	0.131
DES	0.0811	0	0.0822	0.000
DF	-0.0226	0.235		
Dum_2003	-0.0781	0		
Dum_2004	-0.0939	0		
Dum_2005	-0.0934	0		
Dum_2006	-0.1196	0		
Intercept	-1.8403	0	-1.9235	0.000
Log-likelihood	-22219.02		-22241.9	
Number of Obs.	170,553		170,553	
Pseudo R2	0.0024		0.0014	

****p-value* < 0.001, ** *p-value* < 0.01, * *p-value* < 0.05

Data source: rsssf.com

TABLE 9

Snapshot of primitive data prepared after data manipulation:

Year	Day	HomeID	AwayID	SkillH	SkillA	Minute	GoalH	GoalA	StateH	StateA
2005	1	12	19	0.280702	0.517544	1	0	0	0	0
2005	1	12	19	0.280702	0.517544	2	0	0	0	0
2005	1	12	19	0.280702	0.517544	3	0	0	0	0
2005	1	12	19	0.280702	0.517544	4	0	0	0	0
2005	1	12	19	0.280702	0.517544	5	0	0	0	0
2005	1	12	19	0.280702	0.517544	6	0	0	0	0
2005	1	12	19	0.280702	0.517544	7	0	0	0	0
2005	1	12	19	0.280702	0.517544	8	0	0	0	0
2005	1	12	19	0.280702	0.517544	9	0	0	0	0
2005	1	12	19	0.280702	0.517544	10	0	0	0	0
2005	1	12	19	0.280702	0.517544	11	0	0	0	0
2005	1	12	19	0.280702	0.517544	12	0	0	0	0
2005	1	12	19	0.280702	0.517544	13	0	0	0	0
2005	1	12	19	0.280702	0.517544	14	0	0	0	0
2005	1	12	19	0.280702	0.517544	15	0	0	0	0
2005	1	12	19	0.280702	0.517544	16	0	1	0	0
2005	1	12	19	0.280702	0.517544	17	0	0	-1	1
2005	1	12	19	0.280702	0.517544	18	0	0	-1	1
2005	1	12	19	0.280702	0.517544	19	0	0	-1	1

Notes: We do not distinguish between home and away teams in our econometric analysis, but the nature of data allows us to report values separately for home and away teams. Data source: rsssf.com; Day=MatchDay, StateH=State for Home team, StateA=State for Away team

TABLE 10

Summary statistics: Skill variable for both teams
(home and away) – Italy

Skill Variable					
Min	1st Qu.	Median	Mean	3rd Qu.	Max
0.0203	0.3573	0.4174	0.4325	0.4977	0.8509
Skill Difference					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	0.0092	0.03878	0.06648	0.09372	0.4875

Summary statistics: Skill variable for both teams
(home and away) - Spain

Skill Variable					
Min	1st Qu.	Median	Mean	3rd Qu.	Max
0.2322	0.3837	0.4478	0.4603	0.5215	0.7779
Skill Difference					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	0.0055	0.0307	0.0568	0.0838	0.3947

Note: This table reports descriptive statistics for our skill variable. Data source: rsssf.com; Qu=quartile

Italy 2009/10
www.rsssf.com/tables/ital2010.html

Round 1
[Aug 22]
Bologna 1-1 Fiorentina
[Pablo Daniel Osvaldo 24; Adrian Mutu 64]
Siena 1-2 Milan
[Abdelkader Ghezzal 34; Alexandre "Pato" Rodrigues da Silva 29, 48]
[Aug 23]
Catania 1-2 Sampdoria
[Takayuki Morimoto 38; Giampaolo Pazzini 9, Daniele Gastaldello 90+3]
Genoa 3-2 Roma
[Domenico Criscito 49, Alberto Zapater 69, Giuseppe Biava 83;
Rodrigo Taddei 54, Francesco Totti 64]
Inter 1-1 Bari
[Samuel Eto'o 56pen; Vitali Kutuzov 74]
Juventus 1-0 Chievo
[Vincenzo Iaquinta 11]
Lazio 1-0 Atalanta
[Tommaso Rocchi 22]
Livorno 0-0 Cagliari
Palermo 2-1 Napoli
[Edinson Cavani 44, Fabrizio Miccoli 75pen; Marek Hamsik 73]
Udinese 2-2 Parma
[Antonio Di Natale 45+3pen, 89; Alberto Paloschi 42,
Alessandro Lucarelli 49]

Round 2
[Aug 29]
Bari 0-0 Bologna

Note: This is a snapshot of the data from Italy for 2009-10. Link: <http://rsssf.com/tables/ital2010.html> pen=Penalty.

FIGURE 1

Snapshot of data from *rsssf.com* website

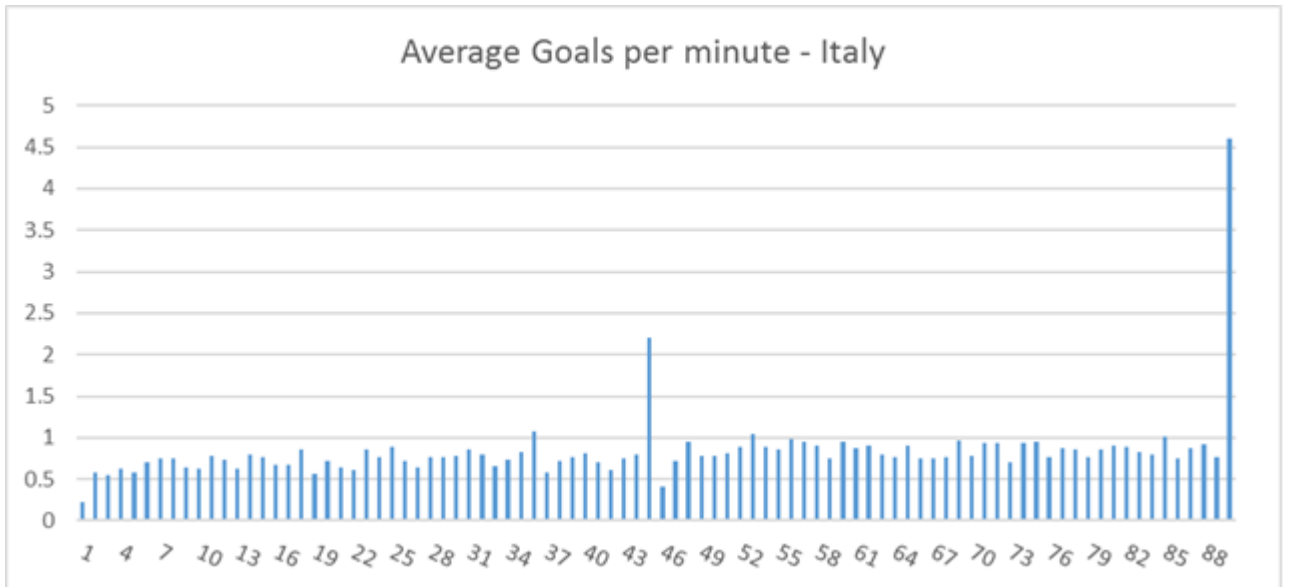


FIGURE 2

Average goals per minute (Italy 2005 to 2011)

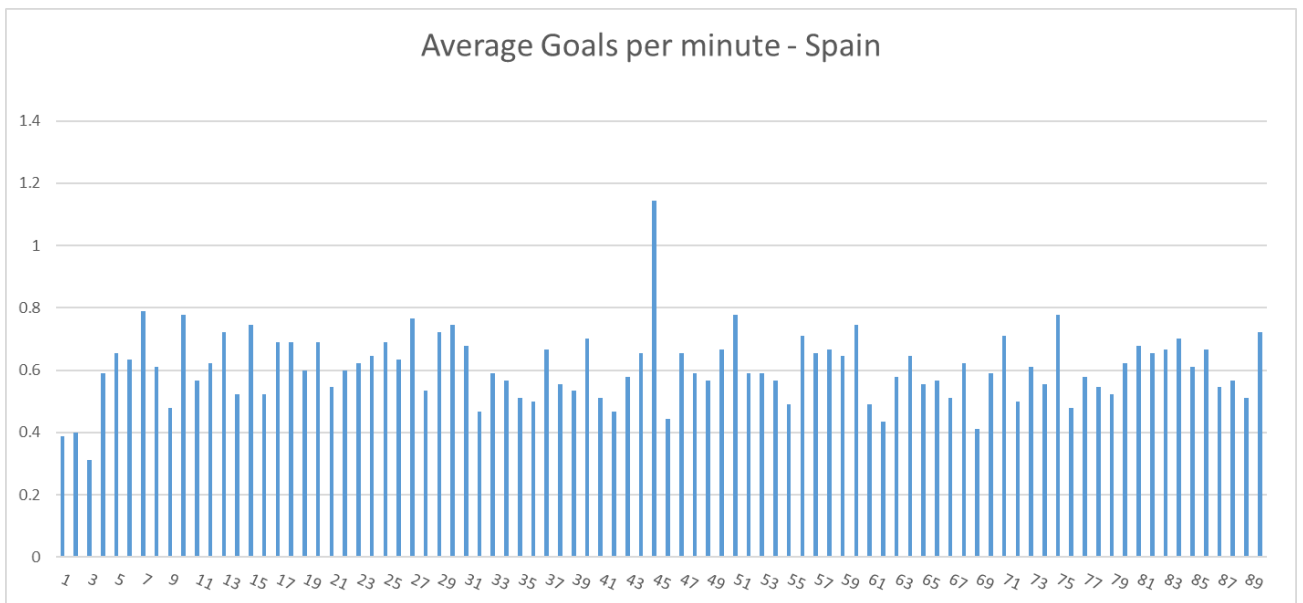


FIGURE 3

Average goals per the minute (Spain 2003 to 2007)

10. References:

- [1] Arrow, K "Higher education as a filter", Journal of Public Economics, 1973
- [2] Amegashie, J, "A nested contest: Tullock meets the All-Pay Auction", 2013
- [3] Atkinson, Stanley and Tschirhart, "Revenue Sharing as an Incentive in an Agency Problem: An Example from the National Football league", The RAND Journal of Economics, Vol. 19, No. 1 (Spring, 1988), pp. 27-43
- [4] Amegashie and Wu, "Self-selection in competing all-pay auctions", 2004
- [5] Brocas, I., and J. D. Carrillo. "Do the Three-Point Victory and Golden Goal Rules Make Soccer More Exciting?" Journal of Sports Economics, 5, 2004, 169-8
- [6] Baye, Kovenock and De Vries, "The solution to the Tullock rent-seeking game when R 2: Mixed-strategy equilibria and mean dissipation rates*", Public Choice 81: 363-380, 1994.
- [7] Baik, H "Effort Levels in Contests with Two Asymmetric Players", Southern Economic Journal, Vol. 61, No. 2 (Oct., 1994), pp. 367-378
- [8] Bell, Murphy and Little, "A Market Share Theorem", Journal of Marketing Research, Vol. 12, No. 2. (May, 1975), pp. 136-141.
- [9] Corchon, C "The theory of contests: a survey", Review of economic design, 2007
- [10] Cornes and Hartley "Dissipation in Rent-seeking contests with entry costs", Keele Economics Research papers, 2002
- [11] Chowdhury and Sheremeta, "A generalized Tullock contest", Public Choice, 2011
- [12] Corral, J., Prieto-Rodríguez, J. and Rob Simmons. "The Effect of Incentives on Sabotage: The Case of Spanish Football." Journal of Sports Economics, 11, 2010, 243-260.
- [13] Dixit, A "The American Economic Review, Vol. 77, No. 5 (Dec., 1987), pp. 891-898
- [14] Dasgupta and Stiglitz "Industrial structure and the nature of Innovation activity" The Economic Journal, 1980
- [15] Dietl, H Egon Franck, Martin Grossmann, Markus Lang, "Contest Theory and its Applications in Sports", 2011
- [16] El-Hodiri and Quirk, "An Economic Model of a Professional Sports League", Journal of Political Economy, Vol. 79, No. 6 (Nov. - Dec., 1971), pp. 1302-1319
- [17] Ferral and Smith, "A Sequential game model of Sports Championship Series: Theory and Estimation", The Review of Economics and Statistics, 1999

- [18] Fearon, J "Rational Explanations for War", International Organization, 1995
- [19] Fu, Q "A theory of affirmative action in college admissions" Economic Inquiry, 2006
- [20] Fort and Quirk, "Cross-subsidization, Incentives and Outcomes in Professional team sports leagues, Journal of Economic Literature, 1995
- [21] Fullerton and McAfee, "Auctioning Entry into Tournaments", Journal of Political Economy, 1999
- [22] Garcia and Rodriguez, "The determinants of football match attendance revisited: Empirical evidence from the Spanish football league."
- [23] Gradstein, Nitzan and Slutsky, "The Effect of Uncertainty on Interactive Behaviour", The Economic Journal, Vol. 102, No. 412 (May, 1992), pp. 554-561
- [24] Goeree and Holt, "Rent-Seeking and the Inefficiency of Non-Market Allocations", Journal of Economic Perspectives, 1999
- [25] Hillman and Katz, "Risk-Averse Rent Seekers and the Social Cost of Monopoly Power", The Economic Journal, Vol. 94, No. 373 (Mar., 1984), pp. 104-110
- [26] Hazlett and Michaels, "The Cost of Rent-Seeking: Evidence from Cellular Telephone License Lotteries", Southern Economic Journal, Vol. 59, No. 3 (Jan., 1993), pp. 425-435
- [27] Hillman and Riley, "Politically contestable rents and transfers" UCLA working papers, 1988
- [28] Haugen, K."Point Score Systems and Competitive Imbalance in Professional Soccer." Journal of Sports Economics, 9, 2008,191-210.
- [29] Konrad, K, "Strategy and dynamics in contests", Oxford University Press, 2009
- [30] Congleton, Hillman, and Konrad, "Forty Years of Research on Rent Seeking: An Overview"
- [31] Konrad and Schlesinger, "Risk Aversion in Rent-Seeking and Rent-Augmenting Games", The Economic Journal, 1997
- [32] Kruger , "Political economy of the rent-seeking soccity", American Economic Review, 1974
- [33] Leach, A "Game, Set and Match: Identifying Incentive response in a tournament", 2003
- [34] Loury "Market Strucure and innovation", Quarterly Journal of Economics, 1979
- [35] Lenininger and Yang "Dynamic rent-seeking contests", Games and Economic behavior, 1994
- [36] Lazear, E "Pay Equality and Industrial Politics", Journal of Political Economy, Vol. 97, No. 3 (Jun., 1989), pp. 561-580
- [37] Moschini, G. "Incentives and Outcomes in a Strategic Setting: The 3- Points-for-a-Win System in Soccer." Economic Inquiry, 48, 2010, 65-79.

- [38] Nitzan, S "Modeling rent-seeking contests", European Journal of political economy, 1994
- [39] Neale, W, "The Peculiar Economics of Professional Sports: A Contribution to the Theory of the Firm in Sporting Competition and in Market Competition", The Quarterly Journal of Economics, Vol. 78, No. 1 (Feb., 1964), pp. 1-14
- [40] Olson M (1965) "The logic of collective action." Harvard University Press, Cambridge
- [41] Palomino, F., L. Rigotti, and A. Rustichini. "Skill, Strategy and Passion: An Empirical Analysis of Soccer." Mimeo, Tilburg University, 2000.
- [42] Palacios-Huerta, I , "Structural changes during a century of the world's most popular sport", Statistical Methods & Applications (2004)
- [43] Rosen, S, "The economics of superstars", American Economic Review, 1981 [44] Rosen, S "Prizes and Incentives in Elimination Tournaments", American Economic Review, 1986
- [45] Rottenberg, Simon "The baseball players labor market", The Journal of Political Economy, 1956
- [46] RKH : Ridder, Kramer, Hopstaken, "Down to Ten: Estimating the Effect of a Red Card in Soccer", Journal of the American Statistical Association, Vol. 89, No. 427 (Sep., 1994), pp. 1124-1127
- [47] Sloane, P "Rottenberg and the Economics of Sport after 50 Years: An Evaluation", IZA Discussion papers, 2006
- [48] Santos, R "Optimal Soccer Strategies.", Economic Inquiry, 52, 2014, pp. 183-200.
- [49] Szymanski, S "The Champions league and the Coase theorem", 2006
- [50] Skaperdas, S "Contest Success Functions", Economic Theory, 1996, pp. 283-290
- [51] Scully, S "Pay and performance in Major League Baseball", American Economic Review, 1979
- [52] Szymanski, S "The economic design of sporting contests" , Journal of economic literature, 2003
- [53] Szymanski and Valletti, "Incentive effects of second prizes", Tanaka Business School Discussion Papers, 2004
- [54] Tullock, G., 1980, Efficient rent-seeking, in: J.M. Buchanan, R.D. Tollison and G. Tullock, Toward a theory of the rent-seeking society (Texas A. & M. University Press, College Station, TX) 97-112. [55] Tollison "Rent-seeking: A survey", 1982
- [56] Vrooman, J "A General Theory of Professional Sports Leagues", Southern Economic Journal, Vol. 61, No. 4 (Apr., 1995), pp. 971-990
- [57] Walker and Wooders, "Minimax Play at Wimbledon", American Economic Review, 2001

CHAPTER 2

HEALTH, TYPES OF LEISURE AND TIME ALLOCATION: EVIDENCE FROM ATUS DATA

1. Introduction

Rising incidents of obesity, adult-onset diabetes and heart disease have led to calls for Americans to devote more time to exercise and other healthful activities. The typical response is that Americans are too busy, and yet work by Aguiar and Hurst (2007) shows that leisure consumption has increased since the 1960s, the period over which the incidence of these adverse health conditions has grown. Meanwhile, better health is presumed to improve productivity and wages at work, both of which would provide an incentive to work more and consume less leisure. Consequently, we may have a puzzle that more leisure is correlated with poorer health, even as leisure time is presumed necessary to invest in exercise and to maintain health. We propose an answer to the puzzle is to distinguish between active and passive forms of leisure.

We propose an empirical procedure to distinguish between the two forms of leisure using a model of time allocation. Results based on the American Time Use Survey (ATUS) show that there are differences in leisure types that correspond to presumed effects on health, and that individuals choose active and passive leisure differently depending on the status of their health. We also consider demographic factors such as age, education, race, location and gender and look at how they affect time allocation choices.

We build a theoretical model where an individual derives utility from consumption of two types of leisure, Active and Passive, and a market good. In our classification, consumption of active leisure involves physical exercise. Active leisure provides utility by the fact that is a current leisure activity⁶ and from expected future utility from improved health. However, also generates disutility from physical exertion which expends a costly resource - calories. Passive leisure only involves current enjoyment of leisure. The theoretical model is designed to identify an empirical strategy that will allow us to categorize leisure activities as either Active or Passive through their interaction with observed health.

⁶ The individual does not earn any income from this activity and it cannot be categorized as home production either. Hence it is a leisure activity which is voluntarily chosen by the individual.

Two strands of literature are relevant to the current paper –

- a) Health being a major determinant of labor market outcomes.
- b) Health being a determinant of time allocation choices

There is a large volume of literature in the first area. Papers in this area are mainly concerning how health may be a determinant of retirement or labor supply. Bound (1991) looks at the merits of using objective measures of health as opposed to self-reported measures in studying retirement models. This paper constructs a statistical model with information from both self-reported and objective measures of health. Rust and Phelan (1997) look at how labor supply of older males are affected by the Social security and Medicare Insurance system. French (2005) also analyzes the effect of health status on retirement by using a model of lifetime decision making. Other notable contributions in this area are by Wu (2003), Coile (2004), Palumbo (2003) and Rosen and Wu (2004). A number of papers have looked at the same issue in Europe (Au et al (2005), Barnay (2010), Currie and Madrian (1999) Disney (2006) and Jones (2010)).

Literature in the second area is surprisingly thin. Biddle and Hamermesh (1990) investigate how time allocated towards sleeping is dependent on various factors. They use aggregated data from 12 countries and show that increases in time spent in the labor market lead to reductions in time allocated towards sleep. They show that higher wage rates reduce time allocated towards sleep for men. This paper shows that time allocated towards sleep responds to economic variables in the same way as other time allocation choices.

Mullahy and Robert (2010) show that although all individuals are endowed with the same time budgets, time allocation choices differ due to heterogeneity in preferences and constraints. They find that educational attainment is positively related to time allocated towards physical activity on weekends or holidays but not during weekdays. They also find that individuals with different human capital endowments allocate time differently to produce health. A recent paper by Podor and Halliday (2012) looks at health status and allocation of time. Their focus is more on how health status affects productivity of home production versus market production. They show that better health is associated with large positive effects on time allocated towards home production and larger positive effects on time spent at market work but less consumption of leisure. Our paper fits into this strand of the literature.

This paper is arranged as follows - Section 2 outlines the theoretical model. Section 3 describes the Econometric model. Section 4 describes the data. Section 5 discusses results of the econometric estimation and describes how various leisure time allocation choices may be aggregated into two broad groups. Section 6 concludes. Tables and figures are at the end of the paper.

2. Theoretical Model

Individuals who are differently endowed with health capital are likely to consume different types of leisure and exhibit different patterns of time use. One of the key features of this paper is the analysis of how allocation of time towards different types of leisure activities is affected by health status of an individual, holding constant other factors such as education, age, gender, location and race.

The theoretical model that we construct helps us carry out this analysis and give us certain testable hypotheses that we proceed to verify using the empirical model. The utility maximization problem for the individual consists of three primary choices – market good, active leisure and passive leisure. The individual derives pleasure from all three. The residual claim on time - market work, is paid hourly remuneration that positively depends on health, $w = w(H)$.

The reasons for making health exogenous in our paper are as follows – we construct a one-period model where individuals come in with a stock of health capital. Health is seen as a stock variable which remains unchanged by actions taken in this period. A limitation of this approach is that we are unable to comment on the effects that consumption of leisure in this period may have on next period's health. However, our main goal is to isolate the effects that health stock may have on various leisure time allocation choices and we find evidence that depending on the nature of the leisure time activity, health may have very different effects.

We have evidence that suggests that the wage earned by an individual in the labor market depends on their stock of health. Respondents with excellent and very good health earn significantly more in the labor market than those with fair and poor health.

With this key insight regarding the correlation between wages and health in mind, we propose a simple one-period model where health is exogenous.

2.1 Individual's problem:

We consider a one-person, one-period model where the individual maximizes utility:

$$\max_{x,A,L} U(x,A,P,H;Z) \quad \text{where}$$

A = Active leisure

P = Passive leisure

T = Time spent working

x = Market good (price = 1)

w = Wage rate

H = Health Stock

Z = Exogenous demographic variables

Subject to two constraints.

- 1) **Budget constraint** which specifies that the individual cannot spend beyond his money income which is obtained by working T hours at the remuneration rate $w(H)$. This income is spent on the market good x , price of which is assumed to be 1.

$$w(H).T = x$$

- 2) **Time constraint:** Total time available is assumed to be 1 without loss of generality. Total time spent at work, on Active and Passive leisure cannot exceed 1.

$$T + P + A = 1$$

Next, we make a key assumption that utility (enjoyment) from Active leisure depends on the level of health. The motivation behind this assumption is simple- since Active leisure requires physical exercise, enjoyment derived from it is related to the health status of the individual. Health capital in a sense 'limits' the amount of Active leisure that can be consumed. We introduce an $f(H)$ term in the utility function which interacts with Active Leisure. This term takes into account the fact that the amount of pleasure that the individual may be able to derive not only depends on their stock capital, but also on their level of efficiency in being able to convert this health stock into consumption of

Active leisure. Thus H enters the utility function indirectly through its effect on Active Leisure.

However, consumption of Passive leisure is not limited by H since such leisure activities do not

require physical exercise and hence P does not interact with health in the utility function described below:

$$U(x, A, P, H; Z) = u(x, A, f(H), P; Z)$$

Where:

$f(H)$ denotes how efficiently health is converted into pleasure derived from Active Leisure. This is likely to increase at a decreasing rate with health. Therefore we assume $f'(H) > 0$, $f''(H) < 0$

Hence, maintaining constraints and assuming health (H) and demographics (Z) to be exogenous, the problem reduces to:

$$\max_{A, L} J \equiv u(w(H).(1 - A - P), A, f(H), P; Z)$$

FOCs:

$$J_A = u_x(w(H).(1 - A - P), f(H).A, P) \cdot (-w(H)) + u_A(w(H).(1 - A - P), f(H)A, P) \cdot f(H) = 0$$

$$J_P = u_x(w(H).(1 - A - P), f(H).A, P) \cdot (-w(H)) + u_P(w(H).(1 - A - P), f(H)A, P) = 0$$

$$\Rightarrow u_A(w(H).(1 - A - P), f(H)A^*, P^*) \cdot f(H) = u_P(w(H).(1 - A - P), f(H)A^*, P^*)$$

Assuming utility function to be separable of the following form:

$$U(x, A, P, H; Z) = u(x; Z) + v(f(H).A, P; Z)$$

FOCs imply that:

$$v_A(f(H)A^*, P^*) \cdot f(H) = v_L(f(H)A^*, P^*)$$

$$(1) \quad f(H) = \frac{v_L(f(H)A^*, P^*)}{v_A(f(H)A^*, P^*)}$$

i.e. the Marginal rate of substitution between active and passive leisure is given by $f(H)$.

The opportunity cost of time is same for both Active and Passive leisure and is equal to the market wage rate. However, wages earned in the labor market are found to respond positively to

health. Keeping this evidence in mind, we assume wages to be an increasing function of exogenous health status for an individual, with $w = w(H)$. However, because of the assumption that Active leisure interacts with health while Passive leisure does not, we obtain result (1) where the marginal rate of substitution between Active and Passive leisure is equal to $f(H)$, where $f(H)$ denotes the efficiency of deriving pleasure from Active leisure. This makes the relationship between w , A , P and H interesting, and I try to examine the relationship between them with our econometric estimation.

Next, I compute comparative statics results to examine the equilibrium effects of health on Active and Passive leisure consumption.

2.2 Comparative Statics

$$J_{AA} \frac{\partial A}{\partial H} + J_{AP} \frac{\partial P}{\partial H} + J_{AH} = 0$$

$$J_{PA} \frac{\partial A}{\partial H} + J_{PP} \frac{\partial P}{\partial H} + J_{PH} = 0$$

So:

$$\begin{bmatrix} J_{AA} & J_{AP} \\ J_{PA} & J_{PP} \end{bmatrix} \begin{bmatrix} \partial A / \partial H \\ \partial P / \partial H \end{bmatrix} = \begin{bmatrix} -J_{AH} \\ -J_{PH} \end{bmatrix}$$

$$\frac{\partial A^*}{\partial H} = \frac{1}{\Delta} \begin{vmatrix} -J_{AH} & J_{AP} \\ -J_{PH} & J_{PP} \end{vmatrix} = \frac{1}{\Delta} (-J_{AH}J_{PP} + J_{AP}J_{PH})$$

$$\frac{\partial P^*}{\partial H} = \frac{1}{\Delta} \begin{vmatrix} J_{AA} & -J_{AH} \\ J_{PA} & -J_{PH} \end{vmatrix} = \frac{1}{\Delta} (-J_{AA}J_{PH} + J_{PA}J_{AH})$$

where $\Delta \equiv J_{AA}J_{PP} - J_{AP}J_{PA}$ (assume $\Delta > 0$, to satisfy Second Order Sufficiency condition), and all derivatives are evaluated at the optimal solution.

Partial effects (economizing on notation):

$$J_{AA} = u_{xx} \cdot w^2 + u_{AA} \cdot f^2$$

$$J_{AP} = u_{xx} \cdot w^2 + u_{AP} \cdot f$$

$$J_{PA} = u_{xx} \cdot w^2 + u_{PA} \cdot f$$

$$J_{PP} = u_{xx} \cdot w^2 + u_{PP}$$

$$J_{AH} = u_{xx} \cdot (-w) \cdot (-w') \cdot (1 - A - P) + v_{AA} \cdot f \cdot f' \cdot A$$

$$J_{PH} = u_{xx} \cdot (-w) \cdot (-w') \cdot (1 - A - P) + v_{PA} \cdot f' \cdot A$$

$$(2A) \quad \frac{\partial A^*}{\partial H} = \left\{ \begin{array}{l} -[u_{xx} \cdot w \cdot w' \cdot (1 - A - P) + v_{AA} \cdot f \cdot f' \cdot A] \cdot [u_{xx} \cdot w^2 + u_{PP}] \\ + [u_{xx} \cdot w^2 + u_{AP} \cdot f] \cdot [u_{xx} \cdot w \cdot w' \cdot (1 - A - P) + v_{PA} \cdot f' \cdot A] \end{array} \right\} \cdot \frac{1}{\Delta}$$

$$(2B) \quad \frac{\partial P^*}{\partial H} = \left\{ \begin{array}{l} -[u_{xx} \cdot w \cdot w' \cdot (1 - A - P) + v_{PA} \cdot f' \cdot A] \cdot [u_{xx} \cdot w^2 + u_{AA} \cdot f^2] \\ + [u_{xx} \cdot w^2 + u_{AP} \cdot f] \cdot [u_{xx} \cdot w \cdot w' \cdot (1 - A - P) + v_{AA} \cdot f \cdot f' \cdot A] \end{array} \right\} \cdot \frac{1}{\Delta}$$

Therefore, the comparative statics results reflect many factors, including the fact that the health stock changes the productivity of work. By inspection of the reduced form solutions (2A) and (2B),

we can see that $\frac{\partial A^*}{\partial H} = A(H, Z) \neq \frac{\partial P^*}{\partial H} = P(H, Z)$. Health capital will have different reduced-form effects on Active and Passive Leisure. In fact, it is possible that the reduced form effects will differ in sign because neither (2A) nor (2B) can be signed. The differences in the reduced form effects of H on Active and Passive leisure provides the key to distinguishing between the two types of leisure empirically and will allow us to aggregate types of leisure into distinct groups that will resemble our theoretical presumptions. As we will see, the two aggregate groups behave in ways consistent with our presumptions about Active and Passive leisure⁷.

⁷ One possible simplification is to treat the amount of time spent at work as fixed and then focus on the relative amounts of Active and Passive leisure. However, we do not adopt this approach since it is inconsistent with the presumption that wages affect leisure demand and hence labor supply. Moreover, fixing labor supply artificially constrains the magnitudes of the active and passive leisure responses to wages in ways that are unlikely to be consistent with the data.

3. Econometric Model

3.1 Time Allocation Choice OLS Estimation:

The American Time Use Survey's Eating and Health (EH) module reports 5 levels of health. Respondents are asked whether they believe that their health falls in one of the following five categories – Excellent, Very Good, Good, Fair or Poor. We are interested in how different levels of health, *ceteris paribus*, affect time allocation choices. Our basic econometric model uses the number of minutes spent on a particular activity as the dependent variable. For each individual i in year t , total leisure time (L_{it}) is decomposed into active (A_{it}) and passive (P_{it}) types according to:

$$L_{it} = P_{it} + A_{it}$$

The estimation is motivated by the theoretical result that the reduced forms $A^* = A(H, Z)$ and $P^* = P(H, Z)$ are different.

The econometric model is –

$$(3) \quad L_{it}^j = \alpha_j + H_{it}'\beta^j + Z_{it}'\gamma^j + \varepsilon_{it}^j$$

Where each leisure type j is denoted by L_{it}^j .

Also, $H = (H_Exc, H_VG, H_Fair, H_Poor)$ denotes the vector of four levels of health that we consider, excluding good health;

$Z = (Edy, Edy^2, Age, Age^2, Female, Metro, White, Black, Dum_2006, Dum_2007)$ denotes the vector of demographic variables.

$\beta = (\beta_{H_Exc}, \beta_{H_VG}, \beta_{H_Fair}, \beta_{H_Poor})^T$ and

$\gamma = (\beta_{Edy}, \beta_{Edy^2}, \beta_{Age}, \beta_{Age^2}, \beta_{Metro}, \beta_{White}, \beta_{Black}, \beta_{Dum_2006}, \beta_{Dum_2007})^T$ denote vectors of coefficients.

j =Leisure type, t =2006 to 2008, Also, H_Exc =Excellent Health, H_VG =Very Good Health, H_Fair =Fair Health, H_Poor =Poor Health⁸

L_{it}^j is the number of minutes spent on j -th leisure activity in year t ; $t=2006, 2007$ and 2008 .

⁸ See 'Data' section for data definitions of other variables

Equation (3) is also used to estimate the effect of health and demographic variables on non-leisure time allocation choices such as work and childcare. We use notations for leisure choices to emphasize the decomposition of total leisure into Active and Passive.

Because leisure choices involve allocations across close substitutes in time, they have the potential of large outliers at the upper and lower tails of the time distribution. To moderate the impact of the outliers on the parameter values, we convert time allocation variables to their standardized form for the j -th leisure type, where the standardized value of the j -th leisure type is defined as:

$$(4) \quad L_std_{it}^j = \frac{L_{it}^j - mean(L_{it}^j)}{SD(L_{it}^j)} \text{ where } j=\text{leisure type}, i=\text{individual}, t=\text{year}.$$

This also simplifies comparison of coefficients across leisure choices because average minutes of leisure consumption differ greatly across leisure types. All coefficients are now interpretable as implying changes in the standard deviation of leisure time. The dependent variable for time allocation regressions reported in table 2 are all in standardized units.

3.2 Wage function estimation:

We also estimate a wage-earning function of the following type:

$$\ln(w_{it}) = \alpha + \beta H_{it} + \gamma Z_{it} + \varepsilon_{it}$$

where H_{it} and Z_{it} are same as above. w_{it} refers to the hourly wage rate. Estimation results are reported in table 5.

4. Data

We use data from the 2006 – 2008 waves of the American Time Use Survey (ATUS). These surveys elicit responses on the time individuals spend on various activities including time spent at market work, household work such as childcare, cooking or cleaning, nonmarket work such as volunteering, and leisure activities such as recreation or watching television. The ATUS sample is drawn from the Current Population Survey (CPS) and includes residents aged 15 or older living in the United States. The sample excludes active military personnel and individuals living in institutions (e.g. hospitals and prisons). Since we are interested in the effect of health on time allocation choices, the best source of data for this purpose is the ATUS Eating and Health (EH) module which was carried out from 2006 to 2008. In addition to time allocation information, the dataset includes self-reported weight and height of the respondent. However, the other aspect of our study is the actual allocation of time by individuals towards different activities in a diary day. The best source of data for this purpose are the ATUS Activity Summary Files. These files contain information about the total number of minutes each respondent spent doing each activity. The level of detail in this dataset is such that every minute out of a total of 1440 minutes in a day are accounted for. The broad categories in the ATUS activity summary files include - personal care, household activities, caring for household members (childcare, adult care), caring for non-household members, work-related activities, education, leisure (includes socializing and relaxing) and sports. This provides us with detailed information on how individuals choose to allocate their time in a typical day.

For this study we merge the ATUS EH module data with the ATUS activity summary files for 2006-2008 using the unique household identifier. Since only one member was interviewed from each household, this makes sure that the individuals in the EH module and the Activity Summary files can be uniquely identified. The data is a pure cross section and different households are surveyed every year. This sample consists of data on 37,832 individuals - 12,891 from 2006, 12,233 from 2007 and 12,708 individuals from 2008.

For the American Time Use Surveys, individuals are randomly selected from a subset of households that have completed their eighth and final month of interviews for the Current Population Survey (CPS). ATUS respondents are interviewed only once about how they spent their time on the previous day, where they were, and whom they were with. The survey is sponsored by the Bureau of Labor Statistics and is conducted by the U.S. Census Bureau.

4.1 Leisure (Active And Passive)

The Activity Summary files in the ATUS use the ATUS activity coding lexicon which is a 3 tier classification system. There are 17 first-tier categories and each first-tier category contains two additional levels of detail. Each activity for an individual is coded using this 3-level classification system and respondent's activities are assigned 6-digit activity codes using this classification system. The number of minutes an individual spends during the diary day in such activities is reported and they add up to 1440. Due to the level of detail in the data, the dataset contains a lot of zeroes. The 17 first-tier classifications are :- personal care (1) , household activities (2), caring for and helping household members (3)(including childcare and adult care), caring for and helping non-household members (4), work and work-related activities (5), education (6), consumer purchases (7), professional and personal care services (8), household services (9), government services and civic obligations (10), eating and drinking (11), leisure (12)(includes socializing and relaxing), sports (13) (includes exercise and recreation), religious and spiritual activities (14), volunteer activities (15), telephone calls (16) and traveling (17). The following are the time allocation choices that are categorized as leisure in this study. They are all measured in minutes.

Sports: Playing baseball, playing basketball, playing billiards, participation in equestrian sports, fencing, fishing, playing football, golfing, doing gymnastics, playing hockey, participation in martial arts, playing racquet sports, playing rugby, playing soccer, softball, vehicle touring/racing, playing volleyball, walking, participation in water sports, weightlifting/strength training, working out (unspecified), wrestling, and ping pong.

Non-Sports: Doing aerobics, biking, boating, bowling, climbing (includes spelunking and caving), dancing, hiking, hunting, participation in rodeo competitions, rollerblading, running, skiing (includes ice skating and snowboarding), using cardiovascular equipment, doing yoga, bungee jumping.

Socializing: Socializing and communicating with others, attending or hosting social parties/receptions/ceremonies and attending meeting for personal interest.

Television Viewing: Watching television and movies (not religious), television (religious), listening to the radio and listening to/playing music (not radio)

Relaxing: Doing nothing/goofing off/wasting time, hanging around/hanging out (alone), sitting in the hot tub/Jacuzzi/whirlpool/sauna, breaks at work, unspecified activity, watching wife

garden/watching husband cook dinner, lying around/ sitting around, sunbathing, grieving, worrying/crying, watching husband assemble lawnmower, resting/relaxing/lounging, reflecting/daydreaming/fantasizing/wondering, looking at pictures in a photo album or looking at photos on computer or camera.

Arts: Attending performing arts

Tobacco Consumption: Smoking a cigarette/cigar/pipe, smoking marijuana/pot/weed, having a cigarette/rolling a cigarette or chewing tobacco/using recreational drugs.

Games (Indoor): Playing board games/ Scrabble/cards, hitting a piñata, playing games over the Internet, spinning dreidels, hiding matzo/ hiding Easter eggs or working jigsaw puzzle/crossword puzzles.

Computer Use: Unspecified computer use, surfing the internet, downloading files/music/ pictures (personal interest), burning CDs, using social networking or computer programming (personal interest)

Hobbies: Scrapbooking/making a scrapbook, making Halloween costumes (for self), making holiday/ other decorations, dyeing Easter eggs, artistic painting, videotaping/ photography/model making/ jewelry making, making pottery/sculpting/wood working, making Christmas decorations, taking pictures, collecting/organizing stamps or coins, bird watching, researching family tree, reading for personal interest, writing for personal interest.

4.2 Other Time use categories:

Work: The number of hours the individual spends at work during the diary day. The categories included in calculating this variable are as follows – working (main job), working (other jobs) and time spent on other income generating activities.

Sleep: The number of hours of sleep reported by the individual during the diary day. ('t010101' in the Activity Summary files)

Childcare: Caring for and helping household children

4.3 Demographics:

Years of formal education: The ATUS Activity Summary files contain a field called 'PEEDUCA'. It is a categorical variable in the ATUS data. I convert it into numerical values using Appendix table A1.

Female: Dummy indicating the respondent is female.

Age: Age of respondent ('TEAGE' in ATUS dataset)

Number of children: Number of children in the household less than 18 years of age ('TRCHILDNUM' in the Activity Summary files)

White: Dummy variable indicating respondent is White

Black: Dummy variable for Race of respondent being Black (=1 if respondent is black, 0 otherwise) (using the 'PTDTRACE' variable in ATUS)

Other Race: Dummy variable indicating Race other than White or Black.

Metro: Dummy variable indicating respondent lives in a Metropolitan area

Table 1 provides summary statistics for all variables used in this study. As is clear from table 1, the top three categories on which Americans spent most time between 2006 and 2008 are TV watching, socializing and relaxing. The bottom three are Tobacco consumption, Sports and Arts. Since there is large variation in the number of minutes spent on different leisure activities, we convert them into standardized units to make comparison possible. This is explained in detail in the Econometric Model section above. Figure 1 in the appendix illustrates the information in table 1.

5. Results

The econometric model laid out in section 3A is used to estimate the effect of health on different time allocation choices –various leisure activities, work and childcare. Here I first summarize results of the econometric estimation for various leisure types and use insights from them to propose a novel method of aggregating different leisure activities into groups.

5.1 Health vs Time allocation - Effects on leisure:

Table 2 contains results of the econometric estimation with time allocation choices. The dependent variable for all regressions reported in this table is the standardized value of the j -th leisure activity as in equation (4). As mentioned above, the objective of choosing to use the standardized value of the dependent variable is to reduce the effect of outliers in the absolute value of time allocation decisions and to make comparison across leisure-activities possible. A close look at Table 2 reveals a pattern on how health, in particular, affects leisure time allocation choices.

The two types of leisure that undoubtedly require physical exertion are - Sports and Non-Sports. For these two categories, excellent and very good health significantly increase time allocation while fair and poor health significantly reduce time allocation.⁹ Other types of leisure have different relationships to health capital. For example, the most common form of leisure consumed in the American Time Use Survey is Television Viewing. For this leisure type, where individuals spend an average 170 minutes per day in our sample, excellent and very good health have negative effects whereas fair and poor health have significantly positive effects.

For the top 3 time allocation choices in our data are - TV viewing, Socializing & Relaxing. For all three, excellent and very good health have significant negative effects whereas fair and poor health have significant positive effects. Fair health, again, is an exception for Socializing as it has a negative and significant coefficient. For TV viewing, the coefficient for poor health is positive and has a very high value (0.5421). This is much larger than the coefficient for Fair health for TV viewing, which suggests that as health deteriorates, time allocated towards TV viewing goes up significantly. The same holds true for relaxation.

⁹ However, the coefficient for fair health with Non-Sports as the dependent variable, has an unexpected sign - positive and the variable is significant at the 5% level.

Estimation results for time allocated towards Tobacco consumption are interesting, since it follows a unique pattern, not observed for any other leisure activity. Excellent and very good levels of health decrease time spent on tobacco consumption, but Fair and Poor health have virtually no effect.

For Games (indoor) which do not require physical exertion, excellent and very good health both have significant negative effects. For Computer Use and Hobbies, excellent and very good health, respectively, have significant negative effects.

Estimation results for time allocated towards Arts related activities show similar pattern to Sports and Non-Sports i.e. Excellent and Very Good health have significant positive effects.

From these results it is clear that different types of leisure respond very differently to the state of health of the respondent. These results confirm predictions of the theoretical model - that in equilibrium, the effect of health on various leisure activities are different. Hence, we are interested in the question – Can leisure activities be aggregated into Active and Passive groups?

5.2. Aggregation:

Next, we test whether there are significant differences between estimated reduced form relationships between health and leisure time allocation and proceed to group different leisure activities as Active or Passive. Given the econometric model $L_{it}^j = \alpha_j + H_{it}'\beta^j + Z_{it}'\gamma^j + \varepsilon_{it}^j$,¹⁰ we test whether $\beta_j = \beta_{j'}$ for any two leisure types $j \neq j'$. We conduct these tests in a pairwise manner across all possible time allocation pairs. The tests impose 4 restrictions on health because there are four measure of health that we use to estimate time allocation choices. Results of these pairwise tests are reported in table 3. The Null hypothesis is that the four measures of health jointly have the same effect on the two members of the pair. If we cannot reject the null hypothesis of equality, we are able to conclude that the two leisure types are part of the same aggregate leisure group. In contrast, time allocation pairs where we are able to reject the null hypothesis are ones where health does not have similar effects on members of the pair, and the two leisure time allocation choices in question are parts of different aggregate groups since they are inherently different from each other.

¹⁰ Details are in section 3 titled 'Econometric model'

5.3 Active leisure group:

For the F- test between the pair (Sports, Non-Sports), we are unable to reject the Null hypotheses. Hence health status of an individual has similar effects on Sports and Non-Sports, suggesting that they may be in the same aggregate leisure group. Since both require physical exercise, we place them in the aggregate leisure type 'Active'.

Pairwise tests, with Sports as one member of the pair, with other leisure activities suggest all lead to rejection of the null hypothesis. The same is true for Non-Sports. Therefore, none of the other leisure activities can be put in the same aggregate 'Active' leisure group, where Sports and Non-Sports belong.

5.4 Passive leisure group:

Next, we test for differences in response to health status for the 8 other leisure activities. Here results are mixed.

Joint test between the pair (TV-viewing, Tobacco) suggests that they may be put in the same aggregate group. The same is true for joint tests of pairs with TV viewing as one member of the pair and the other member being arts, computer use, relaxation or hobbies, suggesting that these five leisure types may be in the same aggregate group.

The shaded boxes in this table are ones where we are unable to reject the null, suggesting that health has the same effect on the two leisure categories. For the 8 leisure categories except Sports and Non-Sports, we are able not able to reject the null hypotheses in the F-test in most cases, suggesting that these leisure activities are similar to each other and can be put in the same group. The general pattern is that choices that do not require physical exercise may be put in the same aggregate leisure group that we call 'Passive'.

Next we test whether the aggregations based on the disaggregated tests are consistent with two separate groups, Active (Sports and Non-Sports) and Passive (TV watching, Socializing, Tobacco, Arts, Games, Computer Use, Relaxation, Hobbies). The joint F- test of equal reduced form coefficients between the two aggregated groups rejects the null hypotheses that health has the same effect on the two groups. Therefore, not only does the health of an individual have different

effects on different subcategories of leisure time allocation choices, there are substitutabilities that exist between members of the same group which makes aggregation possible. After aggregation, it is confirmed that the two aggregate groups are not similar to one another. Therefore, the key prediction of our theoretical model are confirmed.

To test the validity of such aggregation, we carry out estimation of equation 4 with the standardized value of the aggregate leisure activities AL and PL as the dependent variable. Results are reported in table 6. Results show that excellent and very good health have significant positive effects on AL whereas they have significant negative effects on PL. Fair health has significant positive effects on PL consumption. These results, interpreted jointly with results in table 2 tell a similar story – that better health has positive effects on leisure activities that involve being physically active whereas they have negative effects on those that do not require physical activity which makes aggregation meaningful.

5.5 Wage Function estimation: In this context, wage function estimation results are worth discussing. Table 5 reports estimation results for the wage equation. The earnings function that we estimate is:

$\ln(w_{it}) = \alpha + \beta H_{it} + \gamma Z_{it} + \varepsilon_{it}$, where H_{it} denotes the 4 categories of health and Z_{it} denotes exogenous demographic variables. Estimation results in Table 5 suggest that Excellent and Very Good health are positively related to wages earned in the labor market, whereas Fair and Poor health affect wages negatively. The earnings function estimation results also suggests that a) women earn significantly less wages than men b) individuals living in metropolitan areas earn more than those that do not, c) being African-American has a significant negative effect on wages earned and d) age increases wages earned but at a decreasing rate.

These results, interpreted jointly with results in tables 2 and 3 suggest that health categories which signify good health such as excellent and very good, the effects of which on wages are positive, are also associated with higher consumption of leisure activities which involve being physically active.

Also, health categories such as Fair and Poor, which significantly reduce wages earned, are associated with higher consumption of leisure categories which do not require physical activity.

5.6 Work, Sleep and childcare:

It is useful to compare the two aggregate leisure groups with other time allocation choices. We carry out joint significance tests between leisure activities and time spent at work, sleep and childcare. Here the comparison is done with the aggregate groups (Active and Passive) being one member of the pair being tested and sleep, work or childcare being the other member. Results are reported in table 4.

Biddle and Hamermesh (1990) have developed a model of demand for sleep and assume that individuals derive no utility from sleep and that it has no impact on their market or household productivity. Using these strong assumptions, they show that sleep responds to economic variables and that it is not unusual for people's average daily sleep time to differ by as much as 1 hour at different times in their adult lives. We test if health has the same effect on sleep as on Active or Passive leisure. Our tests reject the null and provide evidence that sleep cannot be considered either to be 'Active' or 'Passive' leisure, and that it is meaningful to treat it as a separate time use category.

Similarly, Childcare is not similar to either AL or PL and hence it cannot be aggregated into either Active or Passive leisure group. Therefore, choosing to treat childcare as a separate time use category makes sense which is consistent with Aguiar and Hurst (2007).

Joint tests of significance regarding effect on health on the time allocation pairs (AL, Work) and (PL, Work) suggest that work is neither similar to Active or Passive leisure. This is not surprising since work produces income and leisure does not.

5.7 Do wages rise with health?

In our theoretical model, we assume that wages respond positively to health. Previously, treating health as exogenous, we investigated its effect on various time allocation choices. Also, the earnings function estimation described in the section above confirms that individuals with Excellent and Very good health earn more than those with fair and poor health. While this provides evidence to support the assumption in the theoretical model that $w = w(H), w' > 0$, it also raises the question – if better health increases earnings, is there any effect of leisure activities, which, themselves are

affected by health differently, on wages? If there is an effect, then do these effects follow the pattern that follows from the two estimations? – that Active leisure, which is consumed more with improvements in health, also raises wages earned whereas Passive Leisure, which is consumed more with deteriorating health reduces wages earned. We carry out this exercise with the necessary caution - that the Leisure Activities are endogenous choices, but their effect on wages may be interesting if results corroborate the story that health affects different types of leisure activities in the same way as it affects wages earned in the labor market.¹¹

To answer this question, we carry out another set of wage function estimations with the aggregated leisure choices – AL and PL on the right hand side. Results are reported in table 7. These estimates are biased because AL and PL are endogenous. However, after using instruments for AL and PL, such an exercise is meaningful if we can learn something about how wages are affected by time allocation choices.

To correct for the endogeneity problem, we adopt a two-stage estimation strategy, where in the first stage the effects of health are estimated on Active and Passive leisure at the aggregate level.¹²In the second stage, estimated values of Active and Passive leisure from the first stage are used as instruments and health is excluded from the right-hand side. Formally, the models that we estimate are:

First stage estimation:

$$AL_{it} = \alpha + \beta_1 H_{it} + \gamma_1 Z_{it} + u_{it}$$

$$PL_{it} = \alpha + \beta_2 H_{it} + \gamma_2 Z_{it} + v_{it}$$

Second stage estimation:

$$\ln(w_{it}) = \alpha + \delta_1 AL_{it} + \delta_2 PL_{it} + \eta Z_{it} + \varepsilon_{it}$$

¹¹ Simple correlation coefficients between wage and leisure choices reported in table A2 suggest that as individuals earn more wages, they consume less leisure for most categories. This is a standard neo-classical result since leisure becomes less attractive as earned income is higher. However, Tobacco and Arts are peculiar categories since time spent on both go up as wages rise. Simple correlation coefficients between the four health categories and Sports and TV viewing in table A3 suggest the same result as the econometric estimation does – that excellent and very good health increase time spent on Sports whereas Poor and Fair health increase TV watching.

¹² These results are reported in table 6 and are discussed in the aggregation section above.

where AL and PL are estimated values of Active and Passive leisure from the first stage regressions with exogenous health. u_i and v_i are zero-mean error terms. Z_{it} refers to the vector of exogenous demographic variables. AL_{it} and PL_{it} are correlated with ε_{it} . 1st stage results are reported in table 6. Second stage results are reported in table 7. Second stage estimations follow the Generalized Method of Moments (GMM) estimation process with robust standard errors.

Results of first stage regressions reiterate the story in estimations carried out for the individual leisure activities, but at the aggregate level – that better health leads to consumption of more Active leisure whereas poor and fair health lead to consumption of more Passive leisure.

As results of the second stage GMM estimation in table 7 shows, when both AL and PL are instrumented and included in the wage estimation, Active leisure is not a significant determinant of wages. However, Passive leisure has a large significant negative effect on wages earned in the labor market. The Stock-Yogo F-test for weak instruments for AL gives a value of 11.66 which suggests that the instruments are weak. Therefore, we carry out the same estimation excluding AL and the large significant negative effect of PL on wages persists.

This brings the story back full circle – Americans have been consuming more leisure over the last 50 years and obesity and other health related illnesses have also been going up. If deteriorating health is correlated with lower wages earned, then just looking at leisure consumption as a whole does not tell the full story. Health affects consumption of different types of leisure in different ways, and one of them (Passive leisure) has a strong negative effect on wages earned.

The puzzle that we are concerned with here is this - leisure consumption has been increasing and health has been deteriorating for Americans, particularly at the lower tail of the distribution, with high levels of obesity related illnesses such as diabetes and heart related problems. However, investments towards health require consumption of leisure that recuperates energy spent at work. Thus, more consumption of leisure should be correlated with better health, but evidence from a landmark study by Aguiar and Hurst (2007) along with rising incidents of health-related incidents suggest that such correlation does not exist. This current study proposes that a solution to this puzzle may be the dichotomy between Active and Passive leisure since Passive Leisure has a large

negative effect on wages earned in the labor market and at the same time it is consumed more as health of an individual in America deteriorates.¹³

In view of the above, the $f(H)$ term in the theoretical model may now be more meaningful— although the opportunity cost of time is the same for Active as well as Passive leisure ($w = w(H)$), since wages respond positively to health, an individual is not indifferent between choosing Active versus Passive leisure as their health improves, because the two types of leisure have different effects on labor market outcomes.

5.8. Demographics vs Time allocation:

In this subsection we summarize the effects of demographic variables on various time allocation choices.

- i. Education – As figure 2 in the appendix suggests, the relationship between education and wages is non-linear. This is why we include the quadratic term for education in the time allocation regressions, anticipating that the relationship between time allocation choices and education may also be non-linear. The estimation results show that the linear term has a negative sign whereas the quadratic term has a positive sign for both Active Leisure subcategories (Sports and Non-Sports). This implies that time allocated towards Active leisure has a U-shaped relationship with education. An inverted U-shaped relationship exists between education and TV viewing.
- ii. Age – A similar story appears to be true for Age. The linear terms are negative and significant for the Active leisure categories and the quadratic term is significant and positive indicating a U-shaped relationship. For TV viewing the relationship with age also appears to be U-shaped. Similar U-shaped relationships exist between age and Arts, Relaxation, Computer Use, Games and Hobbies.
- iii. Gender – There are significant differences in time allocation choices between men and women. Women allocate less time towards Sports and Non-Sports activities. Women also devote less time to TV watching, Relaxation, Tobacco consumption, Games and computer use. Women work less hours than men and devote significantly higher time than men for childcare.

¹³ Further studies should also look at how the distinction between Active and Passive types of work affect consumption of leisure and labor market outcomes, an analysis that I am unable to carry out because the ATUS dataset does not allow me to make this distinction.

- iv. Location – The effects of location on sports is not significant. There are mixed results on other time allocation choices.
- v. Race – Race also does not significantly affect time allocated towards Sports and non-sports.

6. Conclusions

In light of the insights made by Aguiar and Hurst (2007) of an increasing trend in leisure consumption in the US this analysis provides a deeper understanding of leisure consumption from an economist's point of view. We build a theoretical model and carry out empirical analysis using the best available time use dataset (ATUS) today. We are first able to confirm the main prediction of the model – that in equilibrium, health has a different effect on two different types of leisure. We are also able to isolate the effect of health on different time allocation choices – leisure, work and childcare.

Our analysis goes further and is able to identify patterns on how health status of an individual may affect their time allocation choices. On the basis of this analysis we are able to make the claim that activities that involve being physically active can be aggregated into the group 'Active Leisure' where the constituents may be similar to one another and may be substitutes from the consumer's point of view. Other are put in the group 'Passive Leisure'. However, across group comparisons suggest that health has very different effects on consumption of Active vs Passive leisure.

We are also able to comment on how improvements in health may lead to consumption of Active leisure versus Passive leisure. Since better health is correlated with more time spent at work, this leads to lesser consumption of leisure. However, the individual's health fixes the enjoyment that they may derive from Active leisure and hence the marginal rate of substitution between Active and Passive leisure is given by how well they are able to convert health capital into consumption of Active leisure. We conclude that labor market outcomes are tied with health status of an individual which causes them to choose different types of leisure in equilibrium.

This analysis is pertinent to the literature on leisure consumption in the area of labor economics. Further research may look at constructing a multi-period theoretical model where the choice of Active leisure today positively affects health tomorrow and the consumption of passive leisure today harms tomorrow's health. Further research may also make the distinction between Active and Passive work which the ATUS dataset does not allow us to do.

Table 1: Summary Statistics:

Variable	Mean	Standard Deviation
<i>Wages</i>	<i>\$4.08</i>	<i>8.76</i>
<i>Sports</i>	<i>6.08</i>	<i>35.34</i>
<i>Non – Sports</i>	<i>11.42</i>	<i>44.67</i>
<i>Socializing and Communicating</i>	<i>66.66</i>	<i>112.93</i>
<i>Arts and Entertainment</i>	<i>5.77</i>	<i>35.80</i>
<i>TV viewing</i>	<i>170.34</i>	<i>156.23</i>
<i>Tobacco</i>	<i>0.37</i>	<i>4.63</i>
<i>RELAXING and thinking</i>	<i>17.86</i>	<i>64.02</i>
<i>Games</i>	<i>10.66</i>	<i>48.42</i>
<i>Computer Use</i>	<i>8.008</i>	<i>36.77</i>
<i>Hobbies</i>	<i>26.77</i>	<i>67.67</i>
<i>Work</i>	<i>159.1</i>	<i>239.59</i>
<i>Sleep</i>	<i>522.44</i>	<i>138.4</i>
<i>Child Care</i>	<i>29.59</i>	<i>74.6</i>
<i>Age</i>	<i>43.64</i>	<i>14.84</i>
<i>Education</i>	<i>13.62</i>	<i>3.16</i>
<i>Num _ Children</i>	<i>0.93</i>	<i>1.16</i>

Distribution of Health among 37,382 respondents–

- Excellent Health 7,037(18.8%);
- Very Good Health 12,800 (34.2%);
- Good Health 11,191 (29.9%);
- Fair Health 4,652 (12.46%);
- Poor Health 1,648 (4.41%).

Data Source: American Time Use Survey (ATUS) – Eating and health Module from 2006 to 2008

Sample size: 37,328. Time allocation choices are in minutes.

Table 2: Time allocation OLS estimates: (Dependent variable is standardized value of j -th leisure type)

	Sports	Non-Sports	Socializing	TV	Tobacco	Arts	Relaxation	Games	Computer Use	Hobbies	Work	Childcare
(Intercept)	0.9626***	0.3076***	0.5722***	0.2925***	-0.2065**	0.0313	0.5613	0.7881***	0.3232***	-0.3974***	-1.138***	-0.4113***
H_Exc	0.1355***	0.212***	-0.08989***	-0.1868***	-0.0623***	0.0681***	-0.0527***	-0.0548***	-0.0304*	-0.0112	0.0085	0.0355*
H_VG	0.0509***	0.1019***	-0.09612***	-0.1126***	-0.0469***	0.0518***	-0.0501***	-0.0286*	0.0038	-0.0249*	0.02	0.0115
H_Fair	-0.0404*	0.0349*	-0.0966***	0.1889***	-0.006	-0.02148	0.0880***	-0.0069	0.0054	0.0150	-0.1531***	-0.0348*
H_Poor	-0.0633*	-0.0619*	0.0858**	0.5421***	-0.0001	-0.0456.	0.3328***	-0.0255	0.0527*	0.0383	-0.4224***	-0.0612*
Education	-0.0160*	-0.0267***	-0.0300***	0.0122.	0.0258***	0.0041	-0.0509***	0.0296***	0.0227**	0.0335***	0.0174*	0.0118.
Edn_sq	0.0003	0.0011***	0.00036	-0.0020***	-0.0012***	0.00019	0.0010***	-0.0014***	-0.0004	0.00019	0.00025	0.0007**
Age	-0.0302***	-0.0057**	-0.0089***	-0.0050**	0.0055**	-0.0065***	-0.0033.	-0.0333***	-0.0166***	-0.0219***	0.0484***	0.0086***
Age_sq	0.0002***	0.0005**	0.00011***	0.0001***	-0.00006***	0.00004*	0.0001***	0.00032***	0.00013***	0.0003***	-0.00057***	-0.0002***
Female	-0.2307***	-0.1487***	0.02593*	-0.2284***	-0.0309**	-0.0028	-0.0433***	-0.0879***	-0.0856***	0.0880***	-0.2595***	0.2551***
Metro	0.1693***	-0.0397**	-0.0588***	0.0049	-0.0108	0.0203	-0.0804***	-0.0111	0.0332*	0.0476***	-0.0049	0.0184
White	0.0229	0.0014	0.0153	0.0673**	0.0289	0.0261	-0.0542*	-0.0106	-0.0892***	0.1237***	0.0032	-0.0403.
Black	-0.0288	-0.0428	0.1522***	0.3033***	-0.0013	-0.0256	0.19***	-0.0773**	-0.1423***	-0.0609*	0.0152	-0.2231***
Dum_2006	-0.0153	-0.0226.	-0.009897	-0.1845***	0.0337*	0.0037	-0.0378**	-0.0861***	0.0077	-0.1687***	0.1893***	0.149***
Dum_2007	-0.0180	0.0054	-0.02539.	-0.1737***	0.0339*	0.0160	-0.0501***	-0.0785***	-0.0027	-0.1841***	0.2178***	0.1566***
Adjusted R-Squared	0.02378	0.01323	0.01533	0.1156	0.001938	0.004269	0.04237	0.01435	0.007539	0.07335	0.08753	0.07156

*** p -value < 0.001, ** p -value < 0.01, * p -value < 0.05

Number of obs: 37,328. Edn_sq refers to education squared.

Data Source: American Time Use Survey (ATUS) – Eating and health Module from 2006 to 2008

Table 3: Aggregation results based on joint F tests: Null hypothesis: $\beta_i = \beta_j$ ¹⁴

	Sports	Non-Sports	Socializing	TV Viewing	Tobacco	Arts	Games	Computer Use	Relaxation	Hobbies
Sports										
Non-Sports	Cannot Reject Null									
Socializing	Reject Null	Reject Null								
TV Viewing	Reject Null	Reject Null	Reject Null							
Tobacco	Reject Null	Reject Null	Reject Null	Cannot Reject Null						
Arts	Reject Null	Reject Null	Cannot Reject Null	Cannot Reject Null	Cannot Reject Null					
Games	Reject Null	Reject Null	Cannot Reject Null	Reject Null	Reject Null	Reject Null				
Computer Use	Reject Null	Reject Null	Cannot Reject Null	Cannot Reject Null	Cannot Reject Null	Cannot Reject Null	Cannot Reject Null			
Relaxation	Reject Null	Reject Null	Cannot Reject Null	Cannot Reject Null	Cannot Reject Null	Cannot Reject Null	Cannot Reject Null	Cannot Reject Null		
Hobbies	Reject Null	Reject Null	Cannot Reject Null	Cannot Reject Null	Cannot Reject Null	Cannot Reject Null	Cannot Reject Null	Reject Null	Cannot Reject Null	

¹⁴ i and j are time allocation choices. $\beta = (\beta_{H_Exc}, \beta_{H_Good}, \beta_{H_Fair}, \beta_{H_Poor})^T$

Table 4: Results of joint tests of significance (Work, Sleep and Childcare) : Null hypothesis: $\beta_i = \beta_j$

	AL	PL	Work	Sleep
AL				
PL	Reject Null			
Work	Reject Null	Reject Null		
Sleep	Reject Null	Reject Null	Reject Null	
Childcare	Reject Null	Reject Null	Reject Null	

Table 5: Earnings function estimation (Dependent Variable: $\ln(w)$)

	Estimate	Std. Error	t value	Pr(> t)	[95% Conf. Interval]	
H_Exc	0.0489	0.012	4.18	0	.025978	.07182
H_VG	0.0390	0.009	4.12	0	.0204505	.0575
H_Fair	-0.0485	0.014	-3.42	0.001	-0.070	-0.020
H_Poor	-0.0900	0.036	-2.52	0.012	-0.159	-0.0200
Edy	-0.0311	0.008	-4.11	0	-0.045	-0.016
Edy_sq	0.0041	0.000	13.37	0	.0034	.00466
Age	0.0560	0.002	30.66	0	.05243	.0596103
Age_sq	-0.0006	0.000	26.64	0	-0.0006	-0.00053
Dum_2006	0.0339	0.010	3.25	0.001	.0134209	.0543653
Dum_2007	0.0722	0.011	6.83	0	.0514419	.092881
White	-0.0077	0.018	-0.42	0.672	-.0432103	.027860
Black	-0.1031	0.020	-5.04	0	-0.143176	-0.0629
metro	0.0914	0.010	9	0	.0715114	.111340
Female	-0.1896	-0.008	23.7	0	-0.205238	-0.17387
(Intercept)	0.9954	0.065	15.42	0	.8688356	1.12189
Adjusted R-squared: 0.2993						

Sample size: 37,328

Data Source: American Time Use Survey (ATUS) – Eating and health Module from 2006 to 2008

Table 6 - Effect of health on AL and PL		
Dependent Variable	AL	PL
	<i>OLS</i>	<i>OLS</i>
H_Exc	0.211*** (0.034)	-0.1812*** (0.070)
H_VG	0.0487* (0.027)	-0.1183** (0.056)
H_Fair	-0.0426 (0.041)	0.2067** (0.085)
H_Poor	-0.0594 (0.104)	0.2647 (0.214)
Edy	0.0049 (0.022)	0.2057*** (0.045)
Edy_sq	-0.00027 (0.0008)	-0.0083*** (0.001)
Age	-0.0281*** (0.005)	-0.0629*** (0.011)
Age_sq	0.00026*** (0.00006)	0.0008*** (0.0001)
White	0.1230** (0.052)	0.3412** (0.109)
Black	0.0192 (0.059)	0.3827** (0.123)
metro	-0.0686** (0.029)	0.0434 (0.061)
Female	-0.2307*** (0.023)	-0.4071*** (0.048)
Adjusted R-Squared	0.0193	0.0137

Regressions also include controls for annual dummy variables and an intercept term. N=37,328

Data Source – American Time Use Survey

Y variables are standardized values as mentioned in eqn. (4)

Table 7– OLS vs IV estimation for the effect of AL and PL on Log-Wages

Dependent Variable	ln(w)	ln(w)	ln(w)	ln(w)	ln(w)	ln(w)
	<i>OLS</i>	<i>OLS</i>	<i>OLS</i>	<i>IV</i>	<i>IV</i>	<i>IV</i>
AL	0.008*** (0.003)	0.008*** (0.003)		-0.0317 (0.160)	0.2965*** (0.065)	
PL	-0.006*** (0.001)		-0.006*** (0.001)	-0.2912*** (0.112)		-0.2738*** (0.066)
Edy	-0.027*** (0.007)	-0.027*** (0.007)	-0.027*** (0.007)	0.029 (0.027)	-0.0299*** (0.009)	0.025 (0.019)
Edy_sq	0.004*** (0.0003)	0.004*** (0.0003)	0.004*** (0.0003)	0.0016 (0.001)	0.0040*** (0.0003)	0.0017** (0.0008)
Age	0.055*** (0.0018)	0.055*** (0.0018)	0.055*** (0.0018)	0.0367*** (0.011)	0.0641*** (0.003)	0.0387*** (0.005)
Age_sq	-0.0005*** (0.00002)	-0.0005*** (0.00002)	-0.0005*** (0.00002)	-0.00032** (0.0001)	-0.00065*** (0.00003)	-0.0003*** (0.00006)
White	-0.002 (0.018)	-0.002 (0.018)	-0.002 (0.018)	0.0959 (0.065)	-0.0427* (0.025)	0.086** (0.040)
Black	-0.103*** (0.020)	-0.103*** (0.020)	-0.103*** (0.020)	0.0092 (0.061)	-0.1109*** (0.026)	0.0018 (0.046)
metro	0.091*** (0.010)	0.091*** (0.010)	0.091*** (0.010)	0.1018*** (0.021)	0.1111*** (0.014)	0.1031*** (0.019)
Female	-0.191*** (0.008)	-0.191*** (0.008)	-0.191*** (0.008)	-0.3155*** (0.079)	-0.1216*** (0.018)	-0.301*** (0.030)
Adjusted R2	0.2972	0.2963	0.2968	Uncentered 0.8911	Uncentered 0.9527	Uncentered 0.9006
N	12,113	12,113	12,113	12,113	12,113	12,113
Stock Yogo F-test				AL 11.66 PL 5.66	AL 11.66	PL 5.65
Wu Hausman Exogeneity Test				Joint test for AL and PL 28.69 (<0.001) Separately, AL – 0.0035 (0.95) and PL – 15.89 (<0.001)	AL 33.06 (<0.001)	PL 58.85 (<0.001)

All regressions also include controls for annual dummy variables and an intercept term. Data Source – American Time Use Survey

Appendix

Table A1: *Education* variable calculation from ATUS raw data using 'PEEDUCA' field

PEEDUCA Code	Meaning	Education (calculated)
31	Less than 1st grade	0
32	1st, 2nd, 3rd, or 4th grade	2.5
33	5th or 6th grade	5.5
34	7th or 8th grade	7.5
35	9th grade	9
36	10th grade	10
37	11th grade	11
38	12th grade - no diploma	12
39	High school graduate - diploma or equivalent (GED)	12
40	Some college but no degree	12+2=14
41	Associate degree - occupational/vocational	12+4=16
42	Associate degree - academic program	12+4=16
43	Bachelor's degree (BA, AB, BS, etc.)	12+4=16
44	Master's degree (MA, MS, MEng, MEd, MSW, etc.)	12+4+2=18
45	Professional school degree (MD, DDS, DVM, etc.)	12+4=16
46	Doctoral degree (PhD, EdD, etc.)	12+4+2+5=23

Data Source: American Time Use Survey (ATUS) – Eating and health Module from 2006 to 2008

Table A2: Correlation coefficients - Wages and Leisure choices

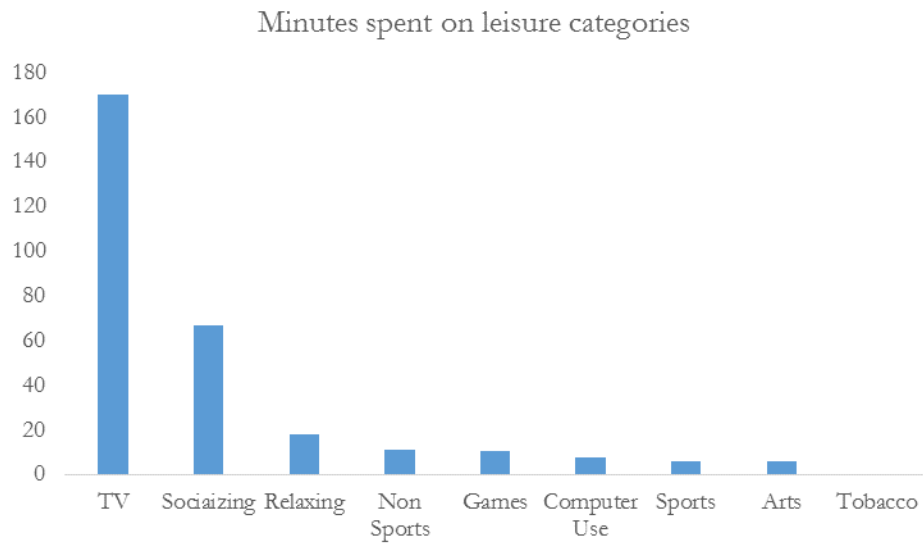
	Wages
Sports	-0.028
Non-Sports	-0.007
TV Viewing	-0.088
Socializing	-0.044
Tobacco	0.0202
Arts	0.0007
Relaxation	-0.034
Games	-0.03
Computer Use	-0.0106
Hobbies	-0.076
Work	0.1874
Sleep	-0.0522

Data Source: American Time Use Survey (ATUS) – Eating and health Module from 2006 to 2008 (Sample size: 37,328)

Table A3: Correlation coefficients: Health and Leisure choices

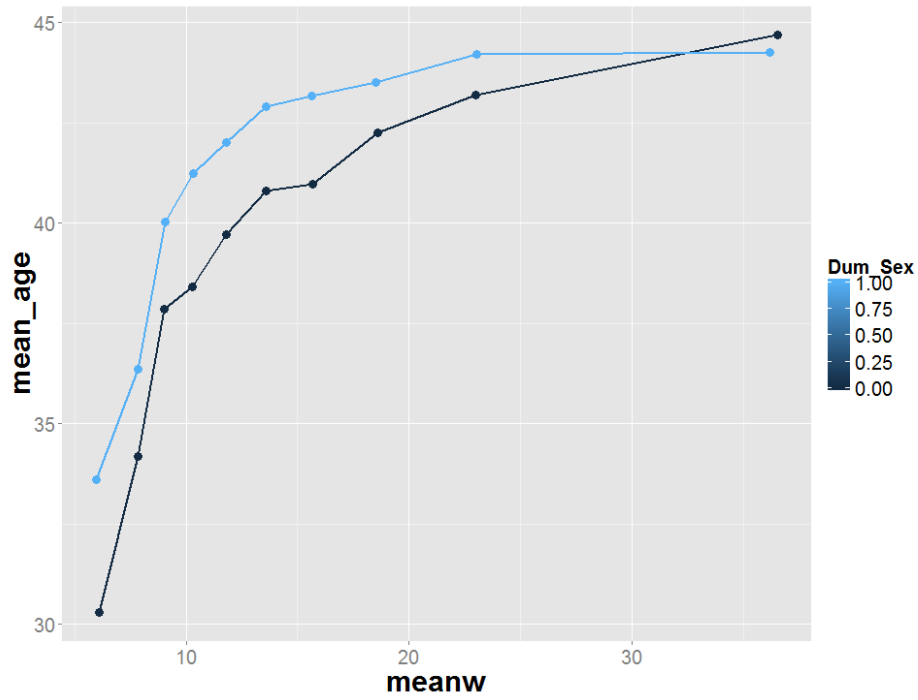
	Sports	TV Viewing
H_Exc	0.0503	-0.1123
H_VG	0.0139	-0.0978
H_Poor	-0.0278	0.016
H_Fair	-0.03334	0.1267

Data Source: American Time Use Survey (ATUS) – Eating and health Module from 2006 to 2008 (Sample size: 37,328)



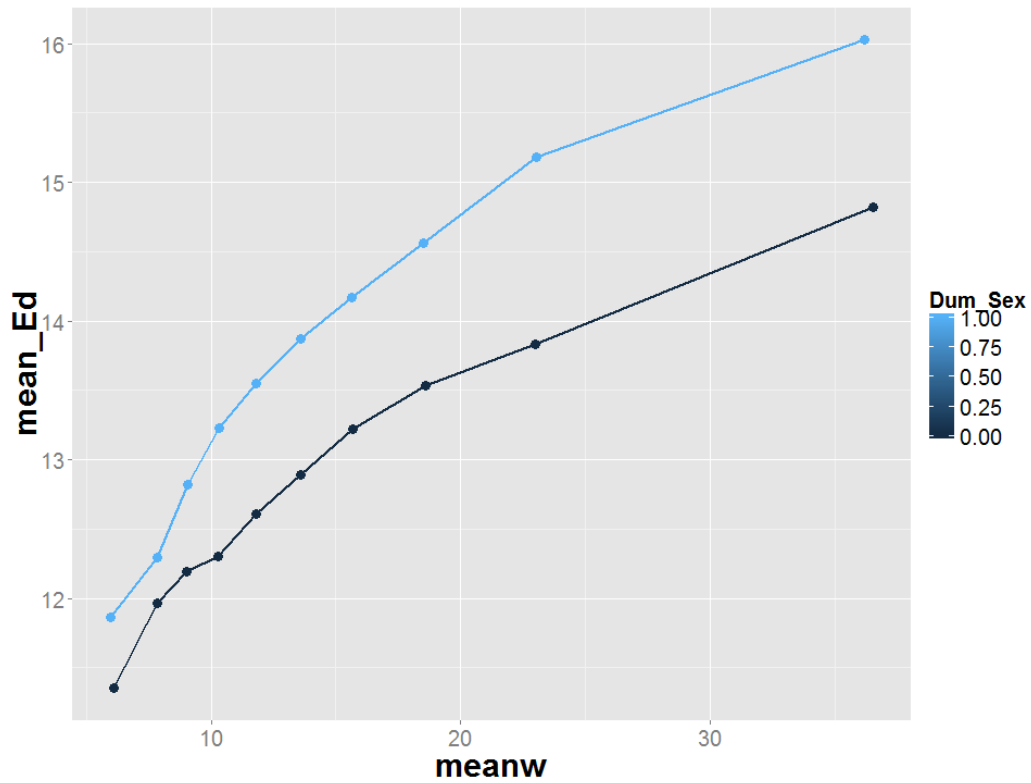
Data Source: American Time Use Survey (ATUS) – Eating and health Module from 2006 to 2008 (Sample size: 37,328)

Figure 1: Number of minutes spent on leisure activities



Data Source: *American Time Use Survey(ATUS) - Activity Summary files and Respondent files from 2006 to 2008*

Figure 2: Mean age versus mean wages (after dividing population into 20 equal sized bins)



Data Source: *American Time Use Survey (ATUS) - Activity Summary files and Respondent files from 2006 to 2008*

Figure 3: Mean education versus mean wages (after dividing population into 20 equal sized bins)

Pairwise F-test description:

The joint test of significance between the pair (Sports, Non_Sports).

$$S = X' \beta^S$$

$$E = X' \beta^E$$

where $S = Sports$, $E = Non - Sports$ and $\beta = (\beta_{H_Exc}, \beta_{H_Good}, \beta_{H_Fair}, \beta_{H_Poor})^T$

Null Hypothesis $H_0 : \beta^S = \beta^E$

Joint tests of significance:

Create dummy variable D_E such that:

$D_E = 1$ if E is dependent variable and 0 otherwise.

Therefore, $L_std_{it}^j = S$ if $D_E = 0$ and $T_std_{it}^j = E$ if $D_E = 1$

Econometric model:

$$L_std_{it}^j = X' \beta^S + D_E \cdot X' (\beta^E - \beta^S)$$

Unrestricted model:

$$L_std_{it}^j \sim H_Exc + H_VG + H_Good + H_Poor + Edy + Edy_sq + Age + Age_sq + Dum_Sex + metro + Race_wh + Race_BL + Dum_2006 + Dum_2007 + D_E * \left(H_Exc + H_VG + H_Good + H_Poor + Edy + Edy_sq + Age + Age_sq + Dum_Sex + metro + Race_wh + Race_BL + Dum_2006 + Dum_2007 \right)$$

Restricted Model: No of restrictions (q) = 4

$$L_std_{it}^j \sim H_Exc + H_VG + H_Good + H_Poor + Edy + Edy_sq + Age + Age_sq + Dum_Sex + metro + Race_wh + Race_BL + Dum_2006 + Dum_2007 + D_E * \left(Edy + Edy_sq + Age + Age_sq + Dum_Sex + metro + Race_wh + Race_BL + Dum_2006 + Dum_2007 \right)$$

Test F-Statistic:

$$F_{q,(n-(k+1))} = \frac{(SSR_r - SSR_{ur}) / q}{SSR_{ur} / (n - (k + 1))}$$

SSR_r = Sum of squared residuals from Restricted model

SSR_{ur} = Sum of squared residuals from Unrestricted model

q = Number of restrictions

n = Number of observations

k = Number of independent variables in the unrestricted model. For our data $q = 4$, $n = 75664$

7 References:

- Aguiar and Hurst – “*Measuring Trends in Leisure: The Allocation of Time over Five Decades*”, QJE 2007
- Au, D., Crossley, T., Schellhorn, M., 2005. “*The effects of health shocks and long-term health on the work activity of older Canadians*”. Health Econ. 14, 999–1018
- Barnay, T., 2010. “*In which ways do unhealthy people older than 50 exit the labour market in France?*”. Eur. J. Health Econ. 11, 127–140.
- Biddle J, Hamermesh D. 1990. “*Sleep and the allocation of time.*” Journal of Political Economy 98(5): 922–043.
- Bound J. 1991. “*Self-reported versus objective measures of health in retirement models.*” Journal of Human Resources 26(1): 106–138.
- Currie, J., Madrian, B., 1999. “*Health, Health Insurance and the Labor Market,*” in Handbook of Labor Economics. In: Ashenfelter, Card (Ed.), Vol 3. Elsevier, Amsterdam.
- Disney, R., Emmerson, C., Wakefield, M., 2006. “*Ill health and retirement in Britain, a panel data-based analysis.*” J. Health Econ. 25, 621–649.
- French E. 2005. “*The effect of health, wealth, and wages on labour supply and retirement behaviour.*” Review of Economic Studies 72(2): 395–427.
- Jones, A., Rice, N., Roberts, 2010. “*Sick of work or too sick to work? Evidence on self-reported health shocks and early retirement from the BHPS.*” Econ. Model. 27, 866–880.
- Mullahy and Robert – “*No time to lose? Time constraints and physical activity*”, NBER working papers 2008
- Palumbo M. 2003. “*Uncertain medical expenses and precautionary saving near the end of the life cycle.*” Review of Economic Studies 66(2): 395–421.

Podor and Halliday - "*Health status and the allocation of time*", Health Economics, 2012

Rosen H, Wu S. 2004. "*Portfolio choice and health status.*" Journal of Financial Economics 72(3): 457–484.

Wu S. 2003. "*The effect of health events on the economic status of married couples.*" Journal of Human Resources 38(1): 219–230.

CHAPTER 3

EFFECTS OF MINIMUM WAGE ON EMPLOYMENT, EARNINGS AND WAGE BILL: EVIDENCE FROM THE RETAIL SECTOR IN USA (1991-2014)

1. Introduction

In theoretical labor economics, the employment effects of an increase in minimum wages are unambiguously negative at the lower end of the skill distribution. The negatively sloped demand curve for labor assumes that an increase in the minimum wage will reduce demand for labor, particularly so for industries that pay minimum wages (restaurants, fast food industry etc.). Perfectly competitive employers will cut employment following a rise in minimum wages (Stigler 1946). Stigler also concludes that minimum wages may be ineffective in reducing poverty and there are other efficient alternatives to minimum wages.

However, the central question regarding the minimum wage debate is concerned with the question of elasticity of labor demand. If demand for labor for a sector which is likely to be affected by minimum wage changes is inelastic, then increases in the minimum wage are likely to make workers better off.¹⁵

In a landmark study, Card and Krueger (1994) evaluated the impact of a rise in New Jersey's minimum wage from \$4.25 to \$5.05 on April 1, 1992. Effects of the higher minimum wage in New Jersey were compared with neighboring regions in eastern Pennsylvania where no such increase in minimum wages occurred. Using a Difference-In-Differences (DID) method, Card and Krueger showed that the increase in minimum wages actually led to increases in employment in New Jersey. There were also no negative effects on the number of McDonald's¹⁶ outlets opened in the state of New Jersey.

A recent study by Dube, Lester and Reich (2010) uses a dataset comprising of contiguous counties in USA to show that minimum wage increases have strong positive effects on earnings while there are no significant effects on employment. They stress on the importance of spatial

¹⁵ Minimum wages have to be binding and also inflation adjusted for this to hold.

¹⁶ In 1991, minimum wage workers made up about half the workforce at McDonald's.

analysis for studying the effects of minimum wages on earnings and employment. They criticize traditional national level studies by pointing out the lack of control for spatial autocorrelation in them and their improper construction of control groups. Minimum wage and employment growth both vary significantly over *time* and *space*. For example, till date a number of Southern States do not have state minimum wages (Alabama, Louisiana, Mississippi, South Carolina and Tennessee). Hence any study that fails to control for this spatial autocorrelation will lead to biased estimates.

However, the study by Dube *et al* (2010) contains some theoretical flaws as mentioned in the section below. In this paper, we recognize the presence of spatial heterogeneity in employment growth and minimum wages in USA but adopt a different approach than Dube *et al* (2010). We employ a Difference-In-Differences (DID) method to investigate the effect of minimum wage changes across state borders on employment, earnings and wage bill. In doing so, we construct a large dataset consisting of contiguous counties in USA from 1991 to 2014. The source of data is the Quarterly Census of Employment and Wages (QCEW) from the Bureau of Labor Statistics (BLS).

The paper is arranged as follow – section II provides a brief literature review, section III provides details of empirical strategy. Section IV describes the data. Section V discusses results. Section VI concludes.

2. Literature Review and Motivation

The Fair Labor Standards Act of 1938 was the first legislation that introduced a national minimum wage in USA. This legislation also introduced the forty-hour work week, guaranteed time-and-a-half rates of payment for overtime and prohibited use of ‘oppressive child labor’ for minors. Under this act, an employer must pay each employee the minimum wage, unless the employee is “engaged in an occupation in which the employee customarily and regularly receives more than \$30 a month in tips”.

As Neumark and Wascher (1992) mention, by the early 1980s, both theoretical and empirical research by economists agreed on the negative effects of increases in minimum wages on employment opportunities for workers who are paid minimum wages or wages that are near the minimum wage. This includes a study conducted for the US president by the Minimum Wage Study Commission of 1981. A major shortcoming of these early studies was the use of only time-series data and no control for geographic differences in the minimum wage.

The question regarding whether minimum wage increases lead to increase in employment or not has been one which has mixed results in the academic literature. There are two distinct methodological approaches that can be observed in this literature.¹⁷

First are the traditional national level studies which use variation in minimum wages across states to estimate employment effects. Notable studies are by Neumark and Wascher (1992, 2007).

Second are case-studies that compare adjoining local areas before and after a minimum wage change (Card and Krueger (1994, 2000); Dube, Naidu and Reich (2007)).

A recent paper by Dube, Lester and Reich (2010) use contiguous counties to identify the effects of minimum wage on earnings and employment in restaurants and other low-wage sectors. They recognize that both national level and case-study approaches fail to account for unobserved heterogeneity in employment growth. They consider all local differences in minimum wages between 1990 and 2006 and find that traditional fixed-effects specifications in national level studies exhibit strong downward bias resulting from the presence of unobserved heterogeneity in employment growth because they fail to account for spatial autocorrelation.

As of 2016, there are only 5 states which do not have a minimum wage. They are – Alabama, Louisiana, Mississippi, South Carolina and Tennessee. Minimum wages range from a high of \$10.00 (California and Massachusetts) to a low of \$5.15 (Georgia and Wyoming). Therefore, there is no doubt that a simple time-series analysis is not sufficient while dealing with minimum wage data which has big variations across geographical regions.

As Dube, Lester and Reich (DLR) mention, contiguous border counties represent good control groups for estimating minimum wage effects since any county is more similar to its cross-state counterpart than any other randomly chosen county. National level studies assume that any one county in the United States is as good a control as another. Therefore, we feel that analysis at the border between contiguous counties is the correct approach while analyzing the employment effects of minimum wage changes over USA.

¹⁷ As Dube, Naidu and Reich (2007) mention, a minimum wage increase will have measurable negative employment effects if (a) the policy is binding, (b) Input substitution possibilities are present and/or (c) product demand is price elastic. In the absence of (b) and (c), minimum wage increases can still increase prices, but not reduce employment substantially.

We create a dataset similar to DLR's dataset with contiguous counties across USA from 1991 to 2014. However, we do not estimate employment levels, but adopt a Difference in Differences type of estimation method similar to Card and Krueger (1994). Our findings suggest that demand for labor in the retail trade sector is elastic and therefore minimum wage changes have significant impact on employment, earnings and the wage bill. We also look at the dynamic effects of minimum wage changes and conclude that minimum wage effects persist over several periods. Our analysis uses data at the county level and clusters standard errors at the state border. The motivation behind clustering is to recognize the importance of spatial heterogeneity in the dataset while not undermining the contiguous county characteristic of the data.

Another motivation for this study is the large number of recent changes to state minimum wage laws, particularly after DLR's period of study (1990 to 2006). My motivation behind this paper is to analyze the effect of these recent changes in minimum wages on employment, earnings per worker and wage bill across 102 county pairs in the United States.

3. Empirical Strategy

In this section, we describe the empirical model used to analyze the effect of differences in minimum wage across contiguous county pairs on employment, earnings per worker and the total wage bill for the retail trade sector between 1991 and 2014 across all states in USA. First we discuss DLR's estimation strategy and later describe in detail the full models that we estimate.

3.1 DLR's estimation strategy and problems:

Dube, Lester and Reich (DLR) construct two distinct samples to carry out the empirical analysis in their paper – one with all counties and one with only contiguous border county-pairs. Here we discuss their estimation strategy and results that they obtain by using the latter sample since it is similar to our sample. Their preferred specification for county i in period t is:

$$(1) \ln y_{it} = \alpha + \eta \ln(MW_{it}) + \delta \ln(y_{it}^{TOT}) + \gamma \ln(pop_{it}) + \phi_i + \tau_{pt} + \varepsilon_{ipt}$$

Where y_{it} =private sector employment, MW_{it} =minimum wage, y_{it}^{TOT} =total private sector employment, pop_{it} = county-level population, ϕ_i =county fixed effect, τ_{pt} = pair-specific time effects.

This estimation strategy is unique since it allows for the time fixed effects to vary across county pairs which takes care of variation in minimum wages across geographical areas in the United States.

However, their estimation strategy is flawed as it contains some methodological problems. The current paper takes care of these flaws and proposes a Difference-in-Differences estimation strategy to make the estimation meaningful.

First, we list the shortcomings of DLR's estimation strategy. To start with, they use levels of employment as the dependent variable to analyze the effects of minimum wages on. The level of employment in different counties may be different due to historical factors that are beyond minimum wages. The difference in the levels of employment over time are a more meaningful unit of analysis. The difference in differences approach is the correct way to carry out this analysis using a contiguous county dataset since such a dataset can be best used to find differences in outcome variables across counties which share a state border. Contiguous counties provide good control groups as mentioned by DLR, but the best use of the dataset is using a DID approach.

Second, while estimating the employment effects of changes in minimum wages on restaurant workers who are paid minimum wages, population is taken to be an exogenous variable. Since migration across the state border can be a direct consequence of a higher minimum wage in an adjoining county, *population* is an endogenous variable. DLR paper suffers from this endogeneity problem. The same problem exists for total *employment*, which is used as an explanatory variable without correcting for endogeneity.

The analysis in this paper is similar to DLR since we also assemble a dataset with contiguous counties. However, adopting a difference in differences approach provides strikingly different results in this paper.

Next, we describe the specifications that we use to test the implications of minimum wage changes on employment, earnings per worker and wage bill in contiguous counties

3.2 Employment

Model A: Panel fixed effects-

Standard errors are clustered at the state border with the assumption that counties that share a state border have correlated error terms.

Model A is the standard panel fixed-effects specification. The dependent variable is the employment level in the subject (neighbor) county. The primary independent variable is the minimum wage in the subject (neighbor) county relative to the base wage for the retail trade sector in that county. Both state-specific and county specific fixed effects are included. Also included is a county-specific linear time trend.

i) Levels:

$$(2) \ln Y_{it}^S = \alpha_S + \beta \ln \left(\frac{MW_{it}^S}{W_{it}^S} \right) + \phi_{iS} + \gamma_S \cdot T_t + \theta \ln(Y_{it-1}^S) + \varepsilon_{it}^S$$

$$(3) \ln Y_{jt}^N = \alpha_N + \beta \ln \left(\frac{MW_{jt}^N}{W_{jt}^N} \right) + \phi_{jN} + \gamma_N \cdot T_t + \theta \ln(Y_{jt-1}^N) + \varepsilon_{jt}^N$$

where:

Y_{it}^S, Y_{jt}^N = Employment in

i, j = border counties

S = subject state

N = neighbor state

t = year

α_S, α_N are state-fixed effects

ϕ_{iS}, ϕ_{jN} are county fixed effects

T_t = Linear time trend

Z_{St}, Z_{Nt} are time varying state factors that may affect the labor market

$\varepsilon_{it}^S, \varepsilon_{jt}^N$ are error-terms

ii) Difference between Subject(S) and Neighbor(N) county in period t :

In the first step, we leverage the contiguous county characteristic of our assembled dataset and take the difference between employment levels in subject county and neighbor county who share a state border.

$$(4) \ln \left(\frac{Y_{it}^S}{Y_{jt}^N} \right) = (\alpha_S - \alpha_N) + \beta \ln \left(\frac{\left(\frac{MW_{it}^S}{W_{it}^S} \right)}{\left(\frac{MW_{jt}^N}{W_{jt}^N} \right)} \right) + \theta \ln \left(\frac{Y_{it-1}^S}{Y_{jt-1}^N} \right) + (\phi_{iS} - \phi_{jN}) + (\gamma_S - \gamma_N) \cdot T + u_{ij,t}$$

where $u_{ij,t}$ refers to the difference in the error terms

iii) Difference over Time between Subject and Neighbor county –DID specification:

Next we difference the difference in employment levels between subject and neighbor counties between consecutive years t and $t-1$.¹⁸ Differencing over time cancels out the state fixed effects terms but not the county-specific fixed effects term, which appears in its difference form in the following regression. The error term now represents difference in error terms between subject and neighbor counties over time.

$$(5) \ln \left(\frac{Y_{it}^S / Y_{jt}^N}{Y_{it-1}^S / Y_{jt-1}^N} \right) = \beta \ln \left(\frac{\left(\frac{MW_{it}^S}{W_{it}^S} \right) / \left(\frac{MW_{jt}^N}{W_{jt}^N} \right)}{\left(\frac{MW_{it-1}^S}{W_{it-1}^S} \right) / \left(\frac{MW_{jt-1}^N}{W_{jt-1}^N} \right)} \right) + \theta \ln \left(\frac{Y_{it-1}^S / Y_{jt-1}^N}{Y_{it-2}^S / Y_{jt-2}^N} \right) + (\gamma_S - \gamma_N) + u_{ij,t,t-1}$$

where $u_{ij,t,t-1}$ refers to the difference in the error terms over time.

Model B: Panel fixed effects with lags & leads:

i) Levels:

$$(6) \ln Y_{it}^S = \alpha_S + \beta_1 \ln \left(\frac{MW_{it}^S}{W_{it}^S} \right) + \beta_2 \ln \left(\frac{MW_{it-p}^S}{W_{it-p}^S} \right) + \beta_3 \ln \left(\frac{MW_{it+q}^S}{W_{it+q}^S} \right) + \theta \ln(Y_{it-1}^S) + \phi_{iS} + \gamma_S \cdot T_t + \varepsilon_{it}^S$$

¹⁸ The range of dataset is from 1991 to 2014.

$$(7) \ln Y_{jt}^N = \alpha_N + \beta_1 \ln \left(\frac{MW_{jt}^N}{W_{jt}^N} \right) + \beta_2 \ln \left(\frac{MW_{jt-p}^N}{W_{jt-p}^N} \right) + \beta_3 \ln \left(\frac{MW_{jt+q}^N}{W_{jt+q}^N} \right) + \theta \ln(Y_{jt-1}^N) + \phi_{jN} + \gamma_N T_t + \varepsilon_{jt}^N$$

Where:

p refers to the duration of time span for lags

q refers to the duration of time span for leads

ii) Difference between Subject(S) and Neighbor(N) county in period t :

$$(8) \ln \left(\frac{Y_{it}^S}{Y_{jt}^N} \right) = (\alpha_S - \alpha_N) + \beta_1 \ln \frac{\left(\frac{MW_{it}^S}{W_{it}^S} \right)}{\left(\frac{MW_{jt}^N}{W_{jt}^N} \right)} + \beta_2 \ln \frac{\left(\frac{MW_{it-p}^S}{W_{it-p}^S} \right)}{\left(\frac{MW_{jt-p}^N}{W_{jt-p}^N} \right)} + \beta_3 \ln \frac{\left(\frac{MW_{it+q}^S}{W_{it+q}^S} \right)}{\left(\frac{MW_{jt+q}^N}{W_{jt+q}^N} \right)} + \theta \ln \left(\frac{Y_{it-1}^S}{Y_{jt-1}^N} \right) + (\phi_{iS} - \phi_{jN}) + (\gamma_S - \gamma_N) T + u_{ij,t}$$

where $u_{ij,t}$ refers to the difference in the error terms

iii) Difference over Time between Subject and Neighbor county –DID specification:

$$(9) \ln \left(\frac{Y_{i,t}^S / Y_{j,t}^N}{Y_{i,t-1}^S / Y_{j,t-1}^N} \right) = \beta_1 \ln \frac{\left(\frac{MW_{i,t}^S}{W_{i,t}^S} \right) / \left(\frac{MW_{j,t}^N}{W_{j,t}^N} \right)}{\left(\frac{MW_{i,t-1}^S}{W_{i,t-1}^S} \right) / \left(\frac{MW_{j,t-1}^N}{W_{j,t-1}^N} \right)} + \beta_2 \ln \frac{\left(\frac{MW_{i,t-p}^S}{W_{i,t-p}^S} \right) / \left(\frac{MW_{j,t-p}^N}{W_{j,t-p}^N} \right)}{\left(\frac{MW_{i,t-p-1}^S}{W_{i,t-p-1}^S} \right) / \left(\frac{MW_{j,t-p-1}^N}{W_{j,t-p-1}^N} \right)} + \beta_3 \ln \frac{\left(\frac{MW_{i,t+q}^S}{W_{i,t+q}^S} \right) / \left(\frac{MW_{j,t+q}^N}{W_{j,t+q}^N} \right)}{\left(\frac{MW_{i,t+q-1}^S}{W_{i,t+q-1}^S} \right) / \left(\frac{MW_{j,t+q-1}^N}{W_{j,t+q-1}^N} \right)} + \theta \ln \left(\frac{Y_{i,t-1}^S / Y_{j,t-1}^N}{Y_{i,t-2}^S / Y_{j,t-2}^N} \right) + (\gamma_S - \gamma_N) + u_{ij,t,t-1}$$

where $u_{ij,t,t-1}$ refers to the difference in the error terms over time.

3.3 Earnings Per Worker:

Earnings/ Worker is defined as Total annual earnings/ Employment

Model A: Panel fixed effects-

i) Levels:

$$(10) \quad \ln \left(\frac{Earnings_{it}^S}{Y_{it}^S} \right) = \alpha_S + \beta \ln \left(\frac{MW_{it}^S}{W_{it}^S} \right) + \phi_{iS} + \gamma_S \cdot T_t + \theta \ln \left(\frac{Earnings_{it-1}^S}{Y_{it-1}^S} \right) + \varepsilon_{it}^S$$

$$\ln \left(\frac{Earnings_{jt}^N}{Y_{jt}^N} \right) = \alpha_N + \beta \ln \left(\frac{MW_{jt}^N}{W_{jt}^N} \right) + \phi_{jN} + \gamma_N \cdot T_t + \theta \ln \left(\frac{Earnings_{jt-1}^N}{Y_{jt-1}^N} \right) + \varepsilon_{jt}^N$$

where:

$Earnings_{it}^S$ = Total earnings in county i

ii) Difference between Subject(S) and Neighbor(N) county in period t :

$$(11) \quad \ln \left(\frac{Earnings_{it}^S}{Y_{it}^S} \right) - \ln \left(\frac{Earnings_{jt}^N}{Y_{jt}^N} \right) = (\alpha_S - \alpha_N) + \beta \ln \left(\frac{MW_{it}^S}{W_{it}^S} \right) - \beta \ln \left(\frac{MW_{jt}^N}{W_{jt}^N} \right) + \theta \ln \left(\frac{Earnings_{it-1}^S}{Y_{it-1}^S} \right) - \theta \ln \left(\frac{Earnings_{jt-1}^N}{Y_{jt-1}^N} \right) + (\phi_{iS} - \phi_{jN}) + (\gamma_S - \gamma_N) \cdot T + u_{ij,t}$$

where $u_{ij,t}$ refers to the difference in the error terms

iii) Difference over Time between Subject and Neighbor county –DID specification:

Next we difference the difference in Earnings per worker levels between subject and neighbor counties between consecutive years t and $t-1$.¹⁹ Differencing over time cancels out the state fixed effects α_S, α_N terms but not the state-specific time trend coefficients γ_S, γ_N which appear in their difference form in the following regression. The error term now represents difference in error terms between subject and neighbor counties over time.

¹⁹ The range of dataset is from 1991 to 2014.

(12)

$$\ln \frac{\left(\frac{Earnings_{it}^S}{Y_{it}^S} \right) / \left(\frac{Earnings_{jt}^N}{Y_{jt}^N} \right)}{\left(\frac{Earnings_{jt-1}^S}{Y_{jt-1}^S} \right) / \left(\frac{Earnings_{jt-1}^N}{Y_{jt-1}^N} \right)} = \beta \ln \frac{\left(\frac{MW_{it}^S}{W_{it}^S} \right) / \left(\frac{MW_{jt}^N}{W_{jt}^N} \right)}{\left(\frac{MW_{it-1}^S}{W_{it-1}^S} \right) / \left(\frac{MW_{jt-1}^N}{W_{jt-1}^N} \right)} + \theta \ln \frac{\left(\frac{Earnings_{it-1}^S}{Y_{it-1}^S} \right) / \left(\frac{Earnings_{jt-1}^N}{Y_{jt-1}^N} \right)}{\left(\frac{Earnings_{jt-2}^S}{Y_{jt-2}^S} \right) / \left(\frac{Earnings_{jt-2}^N}{Y_{jt-2}^N} \right)} + (\gamma_S - \gamma_N) + u_{ij,t,t-1}$$

where $u_{ij,t,t-1}$ refers to the difference in the error terms over time.

Standard errors are clustered at the state border with the assumption that counties that share a state border have correlated error terms.

Model B: Earnings per worker: Panel fixed effects with lags & leads:

i) Levels:

$$(13) \quad \ln \left(\frac{Earnings_{it}^S}{Y_{it}^S} \right) = \alpha_S + \beta_1 \ln \left(\frac{MW_{it}^S}{W_{it}^S} \right) + \beta_2 \ln \left(\frac{MW_{it-p}^S}{W_{it-p}^S} \right) + \beta_3 \ln \left(\frac{MW_{it+q}^S}{W_{it+q}^S} \right) + \theta \ln \left(\frac{Earnings_{it-1}^S}{Y_{it-1}^S} \right) + \phi_{iS} + \gamma_S T_t + \varepsilon_{it}^S$$

$$(14) \quad \ln \left(\frac{Earnings_{jt}^N}{Y_{jt}^N} \right) = \alpha_N + \beta_1 \ln \left(\frac{MW_{jt}^N}{W_{jt}^N} \right) + \beta_2 \ln \left(\frac{MW_{jt-p}^N}{W_{jt-p}^N} \right) + \beta_3 \ln \left(\frac{MW_{jt+q}^N}{W_{jt+q}^N} \right) + \theta \ln \left(\frac{Earnings_{j,t-1}^N}{Y_{j,t-1}^N} \right) + \phi_{jN} + \gamma_N T_t + \varepsilon_{jt}^N$$

Where:

p refers to the duration of time span for lags

q refers to the duration of time span for leads

ii) Difference between Subject(S) and Neighbor(N) county in period t :

$$\begin{aligned}
(15) \quad \ln \left(\frac{\left(\frac{Earnings_{it}^S}{Y_{it}^S} \right)}{\left(\frac{Earnings_{jt}^N}{Y_{jt}^N} \right)} \right) &= (\alpha_S - \alpha_N) + \beta_1 \ln \left(\frac{\left(\frac{MW_{it}^S}{W_{it}^S} \right)}{\left(\frac{MW_{jt}^N}{W_{jt}^N} \right)} \right) + \beta_2 \ln \left(\frac{\left(\frac{MW_{it-p}^S}{W_{it-p}^S} \right)}{\left(\frac{MW_{jt-p}^N}{W_{jt-p}^N} \right)} \right) + \beta_3 \ln \left(\frac{\left(\frac{MW_{it+q}^S}{W_{it+q}^S} \right)}{\left(\frac{MW_{jt+q}^N}{W_{jt+q}^N} \right)} \right) \\
&+ \theta \ln \left(\frac{\left(\frac{Earnings_{it-1}^S}{Y_{it-1}^S} \right)}{\left(\frac{Earnings_{jt-1}^N}{Y_{jt-1}^N} \right)} \right) + (\phi_{iS} - \phi_{jN}) + (\gamma_S - \gamma_N) \cdot T + u_{ij,t}
\end{aligned}$$

where $u_{ij,t}$ refers to the difference in the error terms

iii) Difference over Time between Subject and Neighbor county –DID specification:

$$\begin{aligned}
(16) \quad \ln \frac{\left(\frac{Earnings_{it}^S}{Y_{it}^S} \right) / \left(\frac{Earnings_{jt}^N}{Y_{jt}^N} \right)}{\left(\frac{Earnings_{jt-1}^S}{Y_{jt-1}^S} \right) / \left(\frac{Earnings_{jt-1}^N}{Y_{jt-1}^N} \right)} &= \beta_1 \ln \frac{\left(\frac{MW_{i,t}^S}{W_{i,t}^S} \right) / \left(\frac{MW_{j,t}^N}{W_{j,t}^N} \right)}{\left(\frac{MW_{i,t-1}^S}{W_{i,t-1}^S} \right) / \left(\frac{MW_{j,t-1}^N}{W_{j,t-1}^N} \right)} + \beta_2 \ln \frac{\left(\frac{MW_{i,t-p}^S}{W_{i,t-p}^S} \right) / \left(\frac{MW_{j,t-p}^N}{W_{j,t-p}^N} \right)}{\left(\frac{MW_{i,t-p-1}^S}{W_{i,t-p-1}^S} \right) / \left(\frac{MW_{j,t-p-1}^N}{W_{j,t-p-1}^N} \right)} \\
&+ \beta_3 \ln \frac{\left(\frac{MW_{i,t+q}^S}{W_{i,t+q}^S} \right) / \left(\frac{MW_{j,t+q}^N}{W_{j,t+q}^N} \right)}{\left(\frac{MW_{i,t+q-1}^S}{W_{i,t+q-1}^S} \right) / \left(\frac{MW_{j,t+q-1}^N}{W_{j,t+q-1}^N} \right)} + \theta \ln \frac{\left(\frac{Earnings_{it-1}^S}{Y_{it-1}^S} \right) / \left(\frac{Earnings_{jt-1}^N}{Y_{jt-1}^N} \right)}{\left(\frac{Earnings_{jt-2}^S}{Y_{jt-2}^S} \right) / \left(\frac{Earnings_{jt-2}^N}{Y_{jt-2}^N} \right)} \\
&+ (\gamma_S - \gamma_N) + u_{ij,t,t-1}
\end{aligned}$$

where $u_{ij,t,t-1}$ refers to the difference in the error terms over time.

3.4 Wage Bill:

Model A: Panel fixed effects-

Model A is the standard panel fixed-effects specification. The dependent variable is the employment level in the subject (neighbor) county. The main independent variable is the minimum wage in the subject (neighbor) county relative to the base wage for the retail trade sector in that county. Both state-specific and county specific fixed effects are included. Also included is a county-specific linear time trend.

i) Levels:

$$(17) \ln(WB)_{it}^S = \alpha_S + \beta \ln\left(\frac{MW_{it}^S}{W_{it}^S}\right) + \phi_{iS} + \gamma_S \cdot T_t + \theta \ln(WB_{it-1}^S) + \varepsilon_{it}^S$$

$$(18) \ln(WB)_{jt}^N = \alpha_N + \beta \ln\left(\frac{MW_{jt}^N}{W_{jt}^N}\right) + \phi_{jN} + \gamma_N \cdot T_t + \theta \ln(WB_{jt-1}^N) + \varepsilon_{jt}^N$$

where:

WB refers to Wage bill

ii) Difference between Subject(S) and Neighbor(N) county in period t :

In the first step, we leverage the contiguous county characteristic of our assembled dataset and take the difference between employment levels in subject county and neighbor county who share a state border.

$$(19) \ln\left(\frac{WB_{it}^S}{WB_{jt}^N}\right) = (\alpha_S - \alpha_N) + \beta \ln\left(\frac{\frac{MW_{it}^S}{W_{it}^S}}{\frac{MW_{jt}^N}{W_{jt}^N}}\right) + \theta \ln\left(\frac{WB_{it-1}^S}{WB_{jt-1}^N}\right) + (\phi_{iS} - \phi_{jN}) + (\gamma_S - \gamma_N) \cdot T_t + u_{ij,t}$$

where $u_{ij,t}$ refers to the difference in the error terms

iii) Difference over Time between Subject and Neighbor county –DID specification:

Next we difference the difference in employment levels between subject and neighbor counties between consecutive years t and $t-1$.²⁰ Differencing over time cancels out the state fixed effects α_S, α_N terms but not the state-specific time trend coefficients γ_S, γ_N which appear in their difference form in the following regression. The error term now represents difference in error terms between subject and neighbor counties over time.

$$(20) \ln \frac{(WB_{it}^S / WB_{jt}^N)}{(WB_{it-1}^S / WB_{jt-1}^N)} = \beta \ln \frac{\left(\frac{MW_{it}^S}{W_{it}^S} \right) / \left(\frac{MW_{jt}^N}{W_{jt}^N} \right)}{\left(\frac{MW_{it-1}^S}{W_{it-1}^S} \right) / \left(\frac{MW_{jt-1}^N}{W_{jt-1}^N} \right)} + \theta \ln \left(\frac{WB_{it-1}^S / WB_{jt-1}^N}{WB_{it-2}^S / WB_{jt-2}^N} \right) + (\gamma_S - \gamma_N) + u_{ij,t,t-1}$$

where $u_{ij,t,t-1}$ refers to the difference in the error terms over time.

This is the equation that we estimate. Standard errors are clustered at the state border with the assumption that counties that share a state border have correlated error terms.

Model B: Panel fixed effects with lags & leads:

i) Levels:

$$(21) \ln WB_{it}^S = \alpha_S + \beta_1 \ln \left(\frac{MW_{it}^S}{W_{it}^S} \right) + \beta_2 \ln \left(\frac{MW_{it-p}^S}{W_{it-p}^S} \right) + \beta_3 \ln \left(\frac{MW_{it+q}^S}{W_{it+q}^S} \right) + \theta \ln(WB_{it-1}^S) + \phi_{iS} + \gamma_S \cdot T_t + \varepsilon_{it}^S$$

$$(22) \ln WB_{jt}^N = \alpha_N + \beta_1 \ln \left(\frac{MW_{jt}^N}{W_{jt}^N} \right) + \beta_2 \ln \left(\frac{MW_{jt-p}^N}{W_{jt-p}^N} \right) + \beta_3 \ln \left(\frac{MW_{jt+q}^N}{W_{jt+q}^N} \right) + \theta \ln(WB_{jt-1}^N) + \phi_{jN} + \gamma_N \cdot T_t + \varepsilon_{jt}^N$$

Where:

p refers to the duration of time span for lags

q refers to the duration of time span for leads

ii) Difference between Subject(S) and Neighbor(N) county in period t :

²⁰ The range of dataset is from 1991 to 2014.

$$(23) \quad \ln \left(\frac{WB_{it}^S}{WB_{jt}^N} \right) = (\alpha_S - \alpha_N) + \beta_1 \ln \left(\frac{\left(\frac{MW_{it}^S}{W_{it}^S} \right)}{\left(\frac{MW_{jt}^N}{W_{jt}^N} \right)} \right) + \beta_2 \ln \left(\frac{\left(\frac{MW_{it-p}^S}{W_{it-p}^S} \right)}{\left(\frac{MW_{jt-p}^N}{W_{jt-p}^N} \right)} \right) + \beta_3 \ln \left(\frac{\left(\frac{MW_{it+q}^S}{W_{it+q}^S} \right)}{\left(\frac{MW_{jt+q}^N}{W_{jt+q}^N} \right)} \right) \\ + \theta \ln \left(\frac{WB_{it-1}^S}{WB_{jt-1}^N} \right) + (\phi_{iS} - \phi_{jN}) + (\gamma_S - \gamma_N) \cdot T + u_{ij,t}$$

where $u_{ij,t}$ refers to the difference in the error terms

iii) Difference over Time between Subject and Neighbor county –DID specification:

$$(24) \quad \ln \left(\frac{WB_{i,t}^S / WB_{j,t}^N}{WB_{i,t-1}^S / WB_{j,t-1}^N} \right) = \beta_1 \ln \left(\frac{\left(\frac{MW_{i,t}^S}{W_{i,t}^S} \right) / \left(\frac{MW_{j,t}^N}{W_{j,t}^N} \right)}{\left(\frac{MW_{i,t-1}^S}{W_{i,t-1}^S} \right) / \left(\frac{MW_{j,t-1}^N}{W_{j,t-1}^N} \right)} \right) + \beta_2 \ln \left(\frac{\left(\frac{MW_{i,t-p}^S}{W_{i,t-p}^S} \right) / \left(\frac{MW_{j,t-p}^N}{W_{j,t-p}^N} \right)}{\left(\frac{MW_{i,t-p-1}^S}{W_{i,t-p-1}^S} \right) / \left(\frac{MW_{j,t-p-1}^N}{W_{j,t-p-1}^N} \right)} \right) \\ + \beta_3 \ln \left(\frac{\left(\frac{MW_{i,t+q}^S}{W_{i,t+q}^S} \right) / \left(\frac{MW_{j,t+q}^N}{W_{j,t+q}^N} \right)}{\left(\frac{MW_{i,t+q-1}^S}{W_{i,t+q-1}^S} \right) / \left(\frac{MW_{j,t+q-1}^N}{W_{j,t+q-1}^N} \right)} \right) + \theta \ln \left(\frac{WB_{i,t-1}^S / WB_{j,t-1}^N}{WB_{i,t-2}^S / WB_{j,t-2}^N} \right) + (\gamma_S - \gamma_N) + u_{ij,t,t-1}$$

where $u_{ij,t,t-1}$ refers to the difference in the error terms over time.

4. Data

We use county-level data from the Quarterly Census of Employment and Wages (QCEW) from 1991 to 2014. This data is a result of tabulations of employment and wages of establishments which report to the Unemployment Insurance (UI) programs of the United States, Puerto Rico and U.S Virgin Islands. This represents about 97% of all wage and salaries of civilian employment in the United States.²¹

²¹ Therefore, the data excludes self-employed workers, agricultural workers in small farms, members of armed forces, elected officials, employees of railroads, some domestic workers, student workers and employees of certain small nonprofit organizations.

The QCEW data uses the North American Industry Classification System (NAICS) to report industry data. This paper uses data on the NAICS code 44-45 which includes ‘Retail Trade’. As the Bureau of Labor Statistics website reports – “the Retail Trade sector comprises establishments engaged in retailing merchandise, generally without transformation, and rendering services incidental to the sale of merchandise”. This sector comprises two main types of retailers – store and non-store retailers.²²

Variables used in this study are:

Y_{it}^S, Y_{it}^N = Employment level in the i -th county in the Subject (Neighbor) State. We use the “annual_avg_emplvl” variable in the QCEW dataset which reports the annual average of monthly employment levels for a given year for this variable.

MW_{it}^S = Minimum Wage in the subject state. Obtained from Bureau of Labor Statistics tables.

Earnings/ worker = Calculated by dividing annual_avg_wkly_wages in the QCEW data which measures average weekly wage based on the 12-monthly employment levels and total annual wage levels by Y_{it}

WB = Total_annual_wages in the QCEW data which measures sum of the four quarterly total wage levels for a given year.

We create a *stpr_id* (state pair id) variable which is the same for any two counties that share the same state border.

We use *fips_codes* to identify counties.

5. Results

5.1 Employment:

²² This is taken directly from the www.bls.gov website. The two types of retailers are: a) Store retailers who operate fixed point-of-sale locations. They are located and designed to attract a high volume of walk-in customers. In general, retail stores have extensive displays of merchandise and use mass-media advertising to attract customers. They typically sell merchandise to the general public for personal or household consumption, but some also serve business and institutional clients. In addition to retailing merchandise, some types of store retailers are also engaged in the provision of after-sales services, such as repair and installation. And b) Non- store retailers, like store retailers, are organized to serve the general public, but their retailing methods differ. The establishments of this subsector reach customers and market merchandise with methods, such as the broadcasting of "infomercials," the broadcasting and publishing of direct-response advertising, the publishing of paper and electronic catalogs, door-to-door solicitation, in-home demonstration, selling from portable stalls (street vendors, except food), and distribution through vending machines.

I estimate equation (4) first. Estimation results are reported in table 1. Table 1 also reports estimation results with log-employment growth as the dependent variable. The log-employment growth variable is the same as the left hand side of equation (5). However, we choose not to estimate equation (5) since the minimum wage effects on employment may be attenuated if we estimate this equation. This is primarily driven by the first expression on the right hand side of equation (5) which measures the difference in minimum wage ratios between the subject and neighboring states between periods t and $t-1$. Since minimum wage changes do not occur frequently, this variable is likely to be 0 for most years, whereas the first expression on the right hand side of equation (4) measures the difference in minimum wages between contiguous counties who share a state boundary. Given the nature of our dataset, which is at the county level and measured annually, it is meaningful to estimate equation 4. We follow this convention for estimating effects of minimum wages on earnings per worker and wage bill as well.

The second column of table 1 reports estimation results of equation (4) with robust standard errors clustered at the state border. The coefficient for the minimum wage term on the right hand side, which measures the difference in MW/W between subject and neighbor states is negative and insignificant at the 10% level.

The third column in table 1 reports similar estimation results with the addition of lags and leads of minimum wage ratios between subject and neighbor state. Lags and leads variables for the subject and neighbor states are defined in equations (6) and (7). We use $p = q = 2$ for our estimations. The lag and lead values of minimum wages are included to measure the dynamic effects of minimum wage changes. The dependent variable for estimations of this kind may be interpreted as one which captures the cumulative response of a minimum wage shock. Employment in a particular county for a particular year can be affected by past changes in minimum wages since firms may change employment decisions after minimum wage changes. On the other hand, anticipated future changes in minimum wages may also lead employers to make changes in hiring decisions now. Therefore, both leads and lags should be considered to measure the dynamic effects of minimum wage changes on employment. DLR also estimate lags and leads in minimum wage values in their paper.

Results in column 3 of table 1 are similar to DLR's findings with regard to lags and leads. DLR also find significant negative effects of future increases in minimum wages on employment

levels. In this paper, we are able to confirm this result using a Difference-In-Differences approach. Lagged values for minimum wages, however, do not tell a consistent story.

The fourth and fifth columns of Table 1 report estimation results with log-employment growth as the dependent variable. As is clear, results in column 4 are similar to those in column 2 and results in column 5 are similar to those in column 2. The only difference is in the estimate for the lagged dependent variable. We use the lagged value of the dependent variable as an explanatory variable to control for the current level of employment as past employment levels are likely to affect current level. However, the coefficient of this variable is not meaningful for this study. The results in columns 4 and 5 show that using employment growth as the dependent variable does not make results stronger.

Another key result that we note is that clustering standard errors at the state border do not affect results significantly. We report results for clustered standard errors in tables 1, 2 and 3 but making this distinction does not add much compared to OLS estimates with no clustering. DLR mention that clustering standard errors makes significant difference to their results and they identify the source of this variation to be the presence of spatial heterogeneity in their dataset. Our findings show that this is not necessarily true if a difference in differences method is adopted.

5.2 Earnings per worker

Table 2 reports estimation results using earnings per worker ratio between subject and neighbor county as the dependent variable. Column (2) provides estimation results for equation (11) Across all four models that we estimate, the effect of minimum wage increases on the relative earnings per worker across counties is found to be negative. This, again, is in contrast to DLR's findings who report positive and significant effects of changes in minimum wages on earnings across all their models. This could be a result of model misspecification and endogeneity which are discussed above.

The addition of lags and leads does not add too much to the estimation results for Earnings. Also, similar to employment, growth in earnings per worker between two consecutive years in the QCEW dataset does not provide more insight into the effect of minimum wages on earnings per worker.

5.3 Wage Bill (WB):

Wage Bill for the retail trade sector can be defined as:

$$(25) \text{WB} = H^d \cdot w$$

where: H^d = Labor demand in hours and w = Wage Rate

Effect of a change in Wage on Wage Bill:

$$(26) \frac{d(\text{WB})}{dw} = \frac{d(H^d \cdot w)}{dw} = H^d \left(1 + \frac{w}{H^d} \cdot \frac{\partial H^d}{\partial w}\right) = H^d (1 + \varepsilon^D) \text{ where: } \varepsilon^D \text{ is the elasticity of labor demand}$$

Therefore:

$$\text{sign}\left(\frac{dC}{dw}\right) < 0 \text{ if } \varepsilon^D < -1$$

$$\text{sign}\left(\frac{dC}{dw}\right) > 0 \text{ if } \varepsilon^D > -1$$

$$\text{And } \frac{dC}{dw} = 0 \text{ if } \varepsilon^D = -1$$

For this paper we use data from the retail trade sector across all counties in the United States of America, where workers have historically been paid minimum wages or wages very close to the legally mandated minimum wage. Therefore, the wage rate considered in this paper is $w = w_{\min}$, where w_{\min} denotes the minimum wage. Thus it is clear from the expression above that increases in the minimum wage reduce the wage bill if labor demand for the retail trade sector is elastic.

Estimation results with the dependent variable as the wage bill are reported in table 3. The second column reports estimation results of equation (19). The fifth column presents estimation result of equation (24). Results show that the effect of minimum wage changes on relative wage bill across contiguous counties is negative for all models, which suggests that labor demand for the retail trade sector is elastic. Also, the effect of an increase in the minimum wage last period on this period's wage bill is positive and significant. Here too, growth in wage bill across two periods does not provide more insight.

6. Conclusions

This paper makes important contributions to the debate on minimum wages. Using a contiguous counties dataset we show that the demand for labor in the retail trade sector is elastic. We also consider dynamic effects of minimum wage changes on employment, earnings per worker and wage bill and show that lags and leads of the change in minimum wages across contiguous counties matter implying that effects of changes in minimum wages may be dynamic.

Our results are significantly different in some aspects from previous studies that have used contiguous counties datasets while studying minimum wage effects on employment and on earnings. We conclude that previous studies such as Dube, Lester and Reich (2010) use estimation methods which are flawed and can be improved upon by adopting a DID approach.

Table 1: Effect of minimum wage changes on Log-Relative-Employment and Log-Relative employment growth

Dependent variable	$\ln\left(\frac{Y_{it}^S}{Y_{jt}^N}\right)$	$\ln\left(\frac{Y_{it}^S}{Y_{jt}^N}\right)$	$\ln\left(\frac{Y_{it}^S / Y_{jt}^N}{Y_{i,t-1}^S / Y_{j,t-1}^N}\right)$	$\ln\left(\frac{Y_{it}^S / Y_{jt}^N}{Y_{i,t-1}^S / Y_{j,t-1}^N}\right)$
$\ln\left(\frac{MW_{it}^S / W_{it}^S}{MW_{jt}^N / W_{jt}^N}\right)$	-0.0142 (0.001)	0.152 (0.021)	-0.014 (0.103)	0.152 (0.0021)
$\ln\left(\frac{Y_{i,t-1}^S}{Y_{j,t-1}^N}\right)$	0.9988 (0.0009)	0.997 (0.0009)	-0.001 (0.199)	-0.0028 (0.017)
$\ln\left(\frac{MW_{i,t-1}^S / W_{i,t-1}^S}{MW_{j,t-1}^N / W_{j,t-1}^N}\right)$		-0.165 (0.022)		-0.165 (0.022)
$\ln\left(\frac{MW_{i,t-2}^S / W_{i,t-2}^S}{MW_{j,t-2}^N / W_{j,t-2}^N}\right)$		0.060 (0.017)		0.060 (0.017)
$\ln\left(\frac{MW_{i,t+1}^S / W_{i,t+1}^S}{MW_{j,t+1}^N / W_{j,t+1}^N}\right)$		-0.019 (0.019)		-0.019 (0.019)
$\ln\left(\frac{MW_{i,t+2}^S / W_{i,t+2}^S}{MW_{j,t+2}^N / W_{j,t+2}^N}\right)$		-0.061 (0.014)		-0.061 (0.014)
Intercept	-0.0015	-0.0013	-0.0015	-0.0013
Num. of Obs.	8,862	6,305	8,862	6,305
PSU's	102	102	102	102

OLS estimation results. Robust standard errors are clustered by state border. p-values are reported in parentheses.

S=Subject State, N=Neighbor State, t=1991 to 2014

PSU =Primary sampling unit – refers to the state-pair-id used to identify if two counties in neighboring states share the same border or not.

Table 2: Effect of minimum wage changes on Log-Earnings-per worker

Dependent variable	$\ln\left(\frac{E_{it}^S / Y_{it}^S}{E_{jt}^N / Y_{jt}^N}\right)$	$\ln\left(\frac{E_{it}^S / Y_{it}^S}{E_{jt}^N / Y_{jt}^N}\right)$	$\ln\left(\frac{(E_{it}^S / Y_{it}^S) / (E_{jt}^N / Y_{jt}^N)}{(E_{i,t-1}^S / Y_{i,t-1}^S) / (E_{j,t-1}^N / Y_{j,t-1}^N)}\right)$	$\ln\left(\frac{(E_{it}^S / Y_{it}^S) / (E_{jt}^N / Y_{jt}^N)}{(E_{i,t-1}^S / Y_{i,t-1}^S) / (E_{j,t-1}^N / Y_{j,t-1}^N)}\right)$
$\ln\left(\frac{MW_{it}^S / W_{it}^S}{MW_{jt}^N / W_{jt}^N}\right)$	-0.472 (0.001)	-0.875 (0.007)	-0.472 (0.001)	-0.875 (0.007)
$\ln\left(\frac{E_{i,t-1}^S / Y_{i,t-1}^S}{E_{j,t-1}^N / Y_{j,t-1}^N}\right)$	0.495 (0.010)	0.927 (0.0001)	-0.504 (0.010)	-0.072 (0.0001)
$\ln\left(\frac{MW_{i,t-1}^S / W_{i,t-1}^S}{MW_{j,t-1}^N / W_{j,t-1}^N}\right)$		0.832 (0.010)		0.832 (0.010)
$\ln\left(\frac{MW_{i,t-2}^S / W_{i,t-2}^S}{MW_{j,t-2}^N / W_{j,t-2}^N}\right)$		-0.038 (0.005)		-0.038 (0.005)
$\ln\left(\frac{MW_{i,t+1}^S / W_{i,t+1}^S}{MW_{j,t+1}^N / W_{j,t+1}^N}\right)$		0.052 (0.006)		0.052 (0.006)
$\ln\left(\frac{MW_{i,t+2}^S / W_{i,t+2}^S}{MW_{j,t+2}^N / W_{j,t+2}^N}\right)$		-0.031 (0.004)		-0.031 (0.004)
Intercept	0.0026	0.0011	0.002	0.0011
Num. of Obs.	8,862	6,305	8,862	6,305
PSU's	102	102	102	102

OLS estimation results. Robust standard errors are clustered by state border. p-values are reported in parentheses.

S=Subject State, N=Neighbor State, t=1991 to 2014

PSU =Primary sampling unit, refers to the state-pair-id used to identify if two counties in neighboring states share the same border or not.

Table 3: Effect of minimum wage changes on Log-Wage Bill

Dependent variable	$\ln\left(\frac{WB_{it}^S}{WB_{jt}^N}\right)$	$\ln\left(\frac{WB_{it}^S}{WB_{jt}^N}\right)$	$\ln\left(\frac{WB_{it}^S / WB_{jt}^N}{WB_{i,t-1}^S / WB_{j,t-1}^N}\right)$	$\ln\left(\frac{WB_{it}^S / WB_{jt}^N}{WB_{i,t-1}^S / WB_{j,t-1}^N}\right)$
$\ln\left(\frac{MW_{it}^S / W_{it}^S}{MW_{jt}^N / W_{jt}^N}\right)$	-0.108 (0.022)	-0.725 (0.022)	-0.108 (0.00)	-0.725 (0.022)
$\ln\left(\frac{WB_{i,t-1}^S}{WB_{j,t-1}^N}\right)$	0.991 (0.0009)	0.997 (0.0009)	-0.008 (0.00)	-0.0002 (0.0009)
$\ln\left(\frac{MW_{i,t-1}^S / W_{i,t-1}^S}{MW_{j,t-1}^N / W_{j,t-1}^N}\right)$		0.730 (0.023)		0.730 (0.023)
$\ln\left(\frac{MW_{i,t-2}^S / W_{i,t-2}^S}{MW_{j,t-2}^N / W_{j,t-2}^N}\right)$		0.028 (0.018)		0.028 (0.018)
$\ln\left(\frac{MW_{i,t+1}^S / W_{i,t+1}^S}{MW_{j,t+1}^N / W_{j,t+1}^N}\right)$		0.034 (0.019)		0.034 (0.019)
$\ln\left(\frac{MW_{i,t+2}^S / W_{i,t+2}^S}{MW_{j,t+2}^N / W_{j,t+2}^N}\right)$		-0.091 (0.015)		-0.091 (0.015)
Intercept	-0.001	-0.0002	-0.0012	-0.0002
Num. of Obs.	8,862	6,305	8,862	6,305
PSU's	102	102	102	102

OLS estimation results. Robust standard errors are clustered by state border. p-values are reported in parentheses.

S=Subject State, N=Neighbor State, t=1991 to 2014

PSU =Primary sampling unit, refers to the state-pair-id used to identify if two counties in neighboring states share the same border or not.

7 References:

- Addison, Blackburn, Cotti “*On the robustness of minimum wage effects: geographically-disparate trends and job growth equations*”, *IZA Journal of Labor Economics*, 2015
- Allegretto, Dube and Reich – “*Do Minimum Wages really reduce teen employment? Accounting for heterogeneity and selectivity in state panel data*”, *Industrial Relations: A Journal of Economy and Society*, 2011
- Bertrand, Duflo, Mullainathan – “*How much should we trust difference in difference estimates?*” *Quarterly Journal of Economics*, 2004
- Burkhauser, Richard V, Couch, Wittenburg “*A reassessment of the new economics of the minimum wage*” *Journal of Labor Economics* 2000
- Card and Krueger – “*Minimum wages and employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania*”, *AER* 1994
- Card and Krueger – “*The New Economics of the Minimum Wage*”, Princeton University Press
- Card, David – “*Do minimum wages reduce employment? A case study of California 1987-1989*”, *Industrial and Labor relations review* 1989
- Dube, Lester and Reich – “*Minimum wage effects across state borders: estimates using contiguous counties*”, *The Review of Economics and Statistics* (2010).
- Dube, Naidu and Reich – “*The Economic Effects of a Citywide Minimum wage*”, *Industrial and Labor Relations Review*, 2007
- Hoffman Saul, “*Are the effects of minimum wage increases always small? A reanalysis of Sabia, Burkhauser and Hansen*”, *Industrial and Labor Relations Review*, 2016
- Meer and West “*Effects of the Minimum Wage on Employment Dynamics*”, *The Journal of Human Resources*, Spring 2016
- Neumark and Wascher – “*Employment effects of minimum and sub-minimum wages: Panel Data on State Minimum Wages Laws*”, *Industrial and Labor Relations Review* 1992
- Neumark, David “*The Economic Effects of Minimum Wages: What Might Missouri Expect from Passage of Proposition B?*”

Neumark and Wascher – “*Minimum wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania: Comment?*”, AER 2000

Neumark and Wascher – “*Minimum Wages, the Earned Income Tax Credit and Employment: Evidence from the Post-Welfare Reform Era?*”, NBER working paper

Neumark, Salas and Wascher – “*Revisiting the minimum wage-employment debate: throwing out the baby with the bathwater?*”, Industrial and Labor Relations Review 2014

Sabia, Burkhauser and Hansen – “*Are the Effects of Minimum Wage Increases always small? New Evidence from a case study of New York State?*”, Industrial and Labor Relations Review 2012

Sabia, Burkhauser, Hansen – “*When good measurement goes wrong: New evidence that New York State’s Minimum Wage Reduced Employment.*”, Industrial and Labor Relations Review 2016

Sabia and Burkhauser – “*Minimum Wages and poverty: Will a \$9.50 federal minimum wage really help the working poor?*” Southern Economic Journal 2010

Stigler – “*The Economics of Minimum Wage legislation?*” AER 1946.