STRATEGIC BEHAVIOR IN RISKY COMPETITIVE SETTINGS: SOME EMPIRICAL APPLICATIONS

Juan Sentana

A THESIS SUBMITTED
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF ECONOMICS
UNIVERSITY OF ESSEX

May 2021

Abstract

In the first chapter of this dissertation, I propose a novel and tractable structural model for ascending auctions with both common and private value components in which heterogeneous bidders exhibit loss aversion. Importantly, I find that loss averse bidders bid noticeably lower than risk neutral ones. I also consider a more general framework in which bidders incorporate into their strategies the information of those bidders who are present but decide not to participate after observing the item put up for auction. This results in bidders reducing the aggressiveness of their bids even further. To empirically assess my model, I use data from storage locker auctions in the popular cable TV show *Storage Wars*, finding that the behavior of most of its bidders is consistent with loss aversion. Thus, I document for the first time the presence of loss aversion in actual ascending auctions.

In Chapter 2, I report the results of a (quasi) field experiment in the training grounds of a professional soccer team to check if individuals, when repeatedly facing the same opponents, satisfy the main mixed strategy equilibrium predictions in soccer penalty kicks, a real-life example of strategic play. This is the first time that the implications of mixed strategy equilibria are tested in the field using repeated observations on specific heterogeneous pairs of players, a situation that rarely repeats in real life. In this respect, I also study the effects of the usual practice of treating heterogeneous rivals as if they all came from the same pair because of the lack of repeated observations for specific pairs. In particular, I show that false rejections may arise when heterogeneous pairs are treated as homogeneous and suggest valid aggregate tests that combine statistics from different opponents. My empirical results suggests that the behavior of most soccer players, when repeatedly facing the same opponents, is consistent with equal scoring probabilities across strategies except for the least professional kickers, as well as with serial independence of player's actions. However, I find dependence between the kicker's and goalkeeper's actions. I also find that the least professional goalkeepers tend to replicate each other's actions. In contrast, players do not seem to follow a reinforcement learning model.

In the third chapter, I prove the numerical equivalence for general categorical variables between many seemingly unrelated independent tests. Specifically, I prove that the Pearson's independence test in a contingency table is numerically equivalent to the Lagrange Multiplier test in several popular linear and non-linear regression models: the multivariate linear probability model, the conditional and unconditional multinomial model, the multinomial logit and probit models; as well as the overidentifying restrictions test in GMM. Therefore, different researchers using different econometric procedures will reach exactly the same conclusions if they use any of those tests. Additionally, I show that the asymptotically equivalent Likelihood Ratio tests in the non-linear regression models are numerically identical, and that the heteroskedasticity-robust Wald tests in the multivariate linear probability model and GMM coincide with the Wald test in the conditional multinomial model. All these equivalences also apply to tests of serial independence in a discrete Markov chain, which can be regarded as a time series analogue of the multinomial model. Finally, I use these tests to analyze if professional soccer players follow optimal mixed strategies in penalty kicks. For some players, my empirical results are not consistent with equal scoring probabilities across strategies. In contrast, I find that player's actions are serially independent.

Acknowledgements

I am deeply grateful to several people who have helped me enormously throughout my research. Firstly, I would like to thank my advisors, Abhimanyu Gupta and Simon Weidenholzer for their insightful and detailed comments, fruitful discussions and support. I would also like to thank Dante Amengual, who has read and provided helpful suggestions to improve many parts of this dissertation.

I am also very much indebted to Adolfo Sánchez and Fernando Rodríguez, as well as all the players from AD Alcorcón, because without them I could not have been able to perform my field study.

Participants at the Econometric Society Summer Meeting in Manchester, the 2019 and 2020 Spanish Economic Association Annual Meetings in Alicante and Salamanca, respectively, and the 2020 Econometric Society Winter Meeting in Nottingham made useful comments on earlier drafts of the three chapters.

Financial support from the University of Essex and Fundacion Ramon Areces for carrying out this dissertation is also gratefully acknowledged.

Last, but by no means least, I wish to thank my family, and especially my parents and brother not only for their enormous patience and encouragement, but also because they have always been a source of inspiration and strength.

Contents

1 Loss Aversion in Storage Locker Auctions **1**

1.1 **Introduction** 1

1.2 **Storage Wars** 3

1.2.1 *The main bidders and the auctioneers* 4

1.2.2 *Description of the data* 5

1.3 **The Model with Loss Aversion** 6

1.3.1 *The stochastic setup* 8

1.3.2 *Econometric methodology* 12

1.4 **The Information of Active Non-Bidding Participants** 13

1.4.1 *The model* 13

1.4.2 *Econometric methodology* 16

1.5 **Empirical Application** 16

1.5.1 *Model specification* 17

1.5.2 *Parameter estimates of the baseline model* 18

1.5.3 *Parameter estimates with active non-bidding participants* 19

1.6 **Conclusions** 20

1.7 **Proofs and Auxiliary Results** 22

1.7.1 *Equilibrium proof* 22

1.7.2 *Mean, variances and covariances of values and signals* 26

1.7.3 *Difference between loss aversion and expected utility* 27

1.7.4 *Special cases* 27

1.7.5 *Winner's curse* 29

1.7.6 *Calculating the likelihood of baseline model* 29

1.7.7 *Calculating the likelihood with active non-bidding participants* 33

1.8 **Tables Chapter 1** 36

1.9 **Graphs Chapter 1** 40

2 Heterogeneous Pairs Play Mixed Strategies in the Soccer Field **47**

2.1 **Introduction** 47

2.2 **Penalty Kicks in Football and Game Theory** 50

2.2.1 *The rules* . . . 50
2.2.2 *The formal setting* . . . 51
2.3 **Heterogenous Opponents** . . . 53
2.3.1 *Allowing for heterogeneity of opponents* . . . 57
2.4 **(Quasi) Field Experiment** . . . 58
2.4.1 *Soccer subjects* . . . 58
2.4.2 *Experimental setup* . . . 58
2.4.3 *Descriptive statistics* . . . 59
2.5 **Empirical Analysis** . . . 60
2.5.1 *Test of equal scoring probabilities* . . . 60
2.5.2 *Test for serial independence* . . . 63
2.5.3 *Test for action independence* . . . 66
2.5.4 *Test for sequential independence* . . . 67
2.6 **Reinforcement Learning** . . . 68
2.6.1 *Pair tests* . . . 69
2.6.2 *Tests allowing for heterogeneity of the opponents* . . . 70
2.7 **Conclusions** . . . 71
2.8 **Proofs and Auxiliary Results** . . . 72
2.8.1 *Proof of Proposition 2.1* . . . 72
2.8.2 *Proof of lack of pure strategies* . . . 73
2.8.3 *Existence and uniqueness of equilibrium* . . . 73
2.8.4 *Finite sample behavior under heterogeneity* . . . 75
2.8.5 *Size experiments* . . . 76
2.8.6 *Power experiments* . . . 79
2.8.7 *Multiple testing issues* . . . 83
2.9 **Tables Chapter 2** . . . 84
2.10 **Graphs Chapter 2** . . . 101

3 Tests For Independence Between Categorical Variables **107**
3.1 **Introduction** . . . 107
3.2 **Econometric Methodology** . . . 110
3.2.1 *Pearson's contingency test* . . . 110

3.2.2 *Multivariate regression* 111
3.2.3 *Conditional multinomial model* 113
3.2.4 *Unconditional multinomial model* 114
3.2.5 *Multinomial probit model* 114
3.2.6 *Multinomial logit model* 115
3.2.7 *GMM* 116
3.3 **Practical Applications** 116
3.3.1 *Applications to mixed strategies in soccer penalty kicks* 116
3.3.2 *Other empirical applications* 118
3.4 **Numerical Equivalence Results** 119
3.4.1 Serial independence tests for Markov chains 121
3.5 **Empirical application to penalty kicks** 123
3.5.1 *Test of equal scoring probabilities* 124
3.5.2 *Test for serial independence* 125
3.6 **Conclusions** 126
3.7 **Proofs** 128
3.7.1 *Proof of Proposition 3.1* 128
3.7.2 *Proof of Proposition 3.2* 136
3.7.3 *Proof of Proposition 3.3* 136
3.7.4 *Finite sample distribution* 140
3.7.5 *Relationship Between Test Statistics When $H = 2$ and $H = 3$* 143
3.7.6 *Kickers and Goalkeepers* 144
3.7.7 *F approximations* 144
3.8 **Tables Chapter 3** 146

4 References **152**

List of Figures

1.1 Loss Aversion Utility Function . . . 40
1.2 Bidding Functions . . . 41
1.3 Log-Bid Functions in Multiple Rounds . . . 42
1.4 Log-Bid Functions with Active Non-Bidding Participants . . . 43
1.5 Distribution of Storage Wars Profit/Losses . . . 44
1.6 Log-Bid Functions of Storage Wars Bidders Baseline Model . . . 45
1.7 Log-Bid Functions of Storage Wars Bidders With Active Non-Bidding Participants 46
2.1 Correlation of a "Heterogenous" Player . . . 101
2.2 Rejection Probabilities . . . 102
2.3 p-value Plots of Tests for Independence with 2 Actions Under Heterogeneity . . . 103
2.4 p-value Plots of Test of Equal Scoring Probabilities with 3 Actions . . . 104
2.5 p-value Plots of Serial Independence Tests with 3 Actions . . . 105
2.6 p-value Plots of Tests for Action Independence with 3 Actions . . . 106

List of Tables

1.1 Auctioneer Behavior . . . 36

1.2 Summary Statistics . . . 36

1.3 Bidder's Frequency Participation . . . 37

1.4 Mean Common Value . . . 37

1.5 Maximum Likelihood Estimates Baseline Model . . . 38

1.6 Maximum Likelihood Estimates With Active Non-Bidding Participants . . . 39

2.1 Distribution of Strategies Combinations and Scoring Rates . . . 84

2.2 Test of Equal Scoring Probabilities Cadete A . . . 85

2.3 Test of Equal Scoring Probabilities Cadete C . . . 86

2.4 Test of Equal Scoring Probabilities Juvenil A . . . 87

2.5 Test of Equal Scoring Probabilities with Heterogeneity . . . 88

2.6 Test for Serial Independence of Actions Cadete A . . . 89

2.7 Test for Serial Independence of Actions Cadete C . . . 90

2.8 Test for Serial Independence of Actions Juvenil A . . . 91

2.9 Test for Serial Independence of Actions with Heterogeneity . . . 92

2.10 Test for Action Independence Cadete A . . . 93

2.11 Test for Action Independence Cadete C . . . 94

2.12 Test for Action Independence Juvenil A . . . 95

2.13 Test for Sequential Independence . . . 96

2.14 Reinforced Learning Cadete A . . . 97

2.15 Reinforced Learning Cadete C . . . 98

2.16 Reinforced Learning Juvenil A . . . 99

2.17 Reinforced Learning under Heterogeneity . . . 100

3.1 Numerical and Asymptotic Equivalence Results . . . 108

3.2 Distribution of Strategies Combinations and Scoring Rates . . . 146

3.3 Test for Equality of Scoring Probabilities with 2 Actions . . . 147

3.4 Test for Equality of Scoring Probabilities with 3 Actions . . . 148

3.5 Test for Equality of Scoring Probabilities with 6 Actions . . . 149

3.6 Test for Serial Independence with 2 Actions . . . 150

3.7 Test for Serial Independence with 3 Actions . . . 151

Chapter 1

Loss Aversion in Storage Locker Auctions

1.1 Introduction

The standard framework in most of the empirical and theoretical auction literatures has been expected utility, often with risk neutral bidders. However, Kahneman and Tversky (1979) criticized expected utility because they found that individuals derive their utility from gains and losses relative to some reference point, rather than from absolute levels of wealth as perfectly rational agents do. They presented a new model of decision making under risk known as "Prospect Theory" whose key feature, *loss aversion*, is that individuals are much more sensitive to reductions than to increases in wealth. Given that bidders often suffer losses as well as gains in the auctions they participate, it is important to explore whether loss aversion might better reflect their bidding behavior.

In this chapter, I propose a novel tractable structural ascending auction model that replaces risk neutrality with loss aversion in the well-known framework with symmetric bidders in Milgrom and Weber (1982) and its asymmetric extension in Hong and Shum (2003). Like in standard models, the utility of a bidder depends only on the difference between his own valuation of the object auctioned and his bid, but in this new specification, the bidder is more sensitive to reductions in wealth than to increases of the same magnitude (see Kahneman and Tversky (1979)). In addition, my proposed model allows for both common and private value components in the bidder's valuations, as well as heterogeneous bidder's characteristics. Importantly, I find that, ceteris paribus, loss averse bidders bid substantially lower than risk neutral ones.

I also consider a more general framework in which bidders incorporate into their strategies the information of those bidders who are present but decide not to participate after observing the item put up for auction. In this respect, I find that bidders reduce the aggressiveness of their bids even further as the number of non-bidding participants increases.

To empirically assess my model, I focus on storage locker auctions, which have gained a lot of popularity in recent years, with 155,000 of them taking place each year in the US alone at an average price of \$425.[1] Specifically, I exploit a unique dataset of 254 actual auctions from the first three seasons of the popular cable TV show *Storage Wars*, which follows a core group of individual bidders who take part in storage locker auctions throughout the State of California. As shown in numerous empirical studies (see List (2006), Post et al (2008), Belot et al (2010) and van Dolder et al (2015) for examples), TV shows provide an environment with substantially larger economic incentives than lab experiments. Therefore, analyzing the behavior of bidders in these auctions seems especially relevant.

An interesting unique feature in storage locker auctions is that the contents of the locker are unknown to both the auctioneer and potential buyers before and throughout the entire auction. This situation has the ideal characteristics for bidders to exhibit loss aversion, a feature that often arises when comparing sure outcomes (not participating in the auction) with a risky prospect (participating and making an uncertain positive or negative profit) (see Kahneman and Tversky (1979) for more details).

Empirically, I find that most *Storage Wars* bidders are loss averse in a model in which there are heterogeneous bidder's characteristics. However, the behavior of the most professional bidder is in line with risk neutrality. Not surprisingly, he is the bidder who bids most aggressively.

Additionally, loss aversion persists when bidders incorporate into their strategies the information of those bidders who are present but decide not to participate after observing the locker to be auctioned. Moreover, my findings confirm the empirical relevance of taking into account the presence of non-bidding participants in ascending auctions.

Previous papers have provided experimental evidence of loss aversion in sealed-bid auctions (see for example Lange and Ratan (2010), Banerji and Gupta (2014), Rosato and Tymula (2019) and Eisenhuth and Grunewald (2020) for the case of independent private values and Balzer and Rosato (2020) for interdependent ones). In contrast, there is little work in ascending auctions. An exception is von Wangenheim (2017), who theoretically showed that under independent private values the second-price sealed-bid auction yields strictly higher revenues than the ascending auction when bidders are expectation-based loss averse (see Kőszegi and Rabin (2006) for more details). Therefore, the first chapter of my thesis makes not only a methodological contribution by incorporating loss aversion to a structural ascending auction model with both private and common value components, but also a substantive one by documenting for the first time the

[1] See <https://www.statisticbrain.com/self-storage-industry-statistics/> for more details.

presence of loss aversion in actual ascending auctions.[2]

The rest of this chapter is organized as follows. Section 1.2 describes the TV show in greater detail and provides a summary of the dataset. In Section 1.3, I discuss my proposed structural model of ascending auctions under loss aversion. Then, in Section 1.4 I introduce a framework that incorporates the signals of the bidders present at the auction who decide not to participate. Finally, in Section 5, I discuss the empirical results. This is followed by the conclusions and several appendices where proofs and additional details can be found.

1.2 Storage Wars

The TV show *Storage Wars*, developed by A&E cable network, first aired on December 2010 and soon became the most watched program in the network's history.

Each episode starts with potential bidders gathering outside a storage facility in the State of California. These facilities have the right to put up for auction the contents of the locker when the rent is not paid for three consecutive months. Before bidders are allowed into the storage facility and see the lockers, the auctioneer explains the rules. The auctions are cash only sales, with all sales being final and the winner the highest cash bidder. But more importantly, bidders can only bid on the entire contents of the locker, not on an item-to-item basis.

The lock of the locker is then broken and bidders have exactly five minutes to look around without stepping inside or opening any boxes. During that time, bidders effectively receive a private noisy signal of the unknown contents, and therefore of the valuation of the locker put up for auction. After those five minutes, the auctioneer announces a suggested opening bid for the locker on sale and starts accepting increasingly higher bids from the bidders in the auction.[3] Unlike sealed-bid auctions, there exists "information transparency", in the sense that the identity of all the bidders and their bids are known during the entire auction. The highest bidder at any given moment has the standing bid, which can only be displaced by a higher bid from another bidder. Throughout the auction, every bidder is given the opportunity to outbid the standing bid.[4] Failure to do so results in the end of the auction, with the locker being sold

[2] Some previous papers have looked at other behavioral biases in ascending auctions. Specifically, Dodonova and Khoroshilov (2005, 2009) argue that bidders with independent private values may feel a quasi-endowment effect toward the object for which they are bidding, so that after making an initial bid of $\$x$ followed by a competitor's bid of $\$(x+1)$, they prefer to pay $\$(x+2)$ to keep the object even though they would never buy the auctioned object for this amount when facing a simple buying decision.

[3] In storage locker auctions there are no reserve prices, i.e. the lowest price at which the seller is willing to sell the item, so in principle, the locker could be sold for $1.

[4] There is no predetermined ending time as in eBay. As a consequence, the practice of sniping, i.e. bidding in the very last seconds (see Roth and Ockenfels (2002)), is irrelevant in this auction.

to the winner at a price equal to his bid.

After all the auctions of the day are completed, the winning bidders go through their lockers sorting the "valuable" content from the rest. When they encounter an unusual, potentially very valuable item, bidders consult with experts to find out the actual value of the item.

Although the private valuation might differ from bidder to bidder because they may have different interests, such as collectibles or household items, there is also a clear common component. For example, if a locker contained a standard but very valuable item such as a brand new motorcycle, its value would be very much the same across bidders.

For all those reasons, a model which allows for both common and private values seems adequate to capture the behavior of bidders in these auctions.

1.2.1 *The main bidders and the auctioneers*

The first three seasons of the show follows four main regular bidders throughout the auctions: Dave Hester (a professional buyer who operates his own auction house), Darrell Sheets (a less experienced storage auction bidder who makes his living by selling in swap meets and through his online store), Jarrod Schulz (an even less experienced storage auction bidder who owns a thrift store) and Barry Weiss (a lifelong antiques collector who had never participated in storage auctions before). Additionally, during the auctions there are other bidders present whose identity are not shown publicly, but whose bids are.[5]

The auctioneers on the show are Dan and Laura Dotson, who have run their own business since 1983: American Auctioneers. Their retribution scheme comes from a small percentage of the locker sale they receive from the storage unit company. Therefore, it is in their interest that the locker is sold at a high price.

One of their key roles is to engage bidders. To accomplish this, they have to start the auction by announcing a suggested opening bid low enough to be immediately accepted by one of the bidders. The regression results in Table 1.1 suggest that the opening bid is set taking into account the location and size of the locker, not surprisingly since it is the only available information.

(Table 1.1)

[5] Given that there is no identifying information on those bidders, I treat them as homogeneous when estimating the empirical model in section 1.5.

1.2.2 *Description of the data*

As explained in the introduction, I examine the bidding behavior of *Storage Wars* participants in 254 actual auctions, which covers seasons 1 (59), 2 (103) and 3 (92) of the TV show.[6] The dataset contains the identity of the bidders, including the four main regular ones, the number of regular bidders present at the auction, as well as the total number of bidders bidding per auction (ranging from 2 to 7), the location of the auction, the size of the locker, whether the main regular bidders decide in real time (live) to bid or not after visually inspecting the locker and before the auction starts, the entire bid sequence and the ex-post value of the locker. I have also collected per capita income data of the municipality where the locker is located, as one would expect a priori that richer neighborhoods have more valuable locker contents.

There are three types of lockers in the auctions: small (10×10 ft.) fitting household items from 3 rooms, medium (10×20 ft.) fitting household items from 5 rooms and large (10×30 ft.) fitting household items from 7 rooms.

Table 1.2 offers a basic description of the data.

(Table 1.2)

For each season, it shows the number of times a small, medium and large locker has been auctioned, the average profit each bidder makes, the average ex-post value of the locker auctioned, the average median household income of the municipalities where the lockers are located and the total number of bidders participating per auction.

The most frequent auctions involve 3, 4 and 5 active bidders, with 50, 72 and 63 auctions, respectively. Additionally, there are many more small and medium size lockers auctioned than large ones. However, after running a standard OLS regression, I find that the order in which the lockers are put up for auction each day is independent of the ex-post value of the locker. This result is not surprising given that the value of the locker is unknown to both the auctioneer and bidders before and throughout the entire auction.

Table 1.3 describes the participation rates of the main bidders in *Storage Wars*.

(Table 1.3)

As can be seen, none of the four main bidders has actually participated in all of the auc-

[6] Video clips of each episode are widely available on the Internet, for example, through the A&E website <https://www.aetv.com/shows/storage-wars>.

tions. Jarrod is the bidder who has participated the most, followed by Darrell, Dave and Barry. However, all four of them only coincide 11.42% of the time. Given that the main bidders often publicly indicate whether they will participate in the auction after looking at the locker and before the auctioneer announces the opening bid, I assess whether their actual participation is in line with their claims using a standard independence test (see section 3.2 for details). The results show that the null hypothesis of independence between their actual participation and their claims is massively rejected for all the main bidders with a p-value of 0, confirming that their participation decisions are coherent with their announcements. This fact motivates the extension of the model in section 1.4, in which bidders incorporate into their strategies the information of those bidders who are present but decide not to participate after observing the item put up for auction.

1.3 The Model with Loss Aversion

The theoretical auction model studied in this chapter resembles the Japanese "button" auction in Milgrom and Weber (1982), in which prices raise continuously, bidders keep pressing a button to remain active, and once a bidder drops out, he cannot reenter the auction at a higher price.[7] More formally, consider an auction of a single item with N potentially heterogeneous bidders, indexed $i = 1, \ldots, N$, for whom the value of the item auctioned is V_i. However, at the beginning of the auction, they only observe a private noisy signal X_i of their own valuation V_i.

The auction proceeds in rounds, indexed $k = 0, \ldots, N-2$, in which active bidders submit bids. A new round starts whenever a bidder drops out and bidders are indexed by the round in which they drop out. Thus, bidder N drops out in round 0 at price P_0 and bidder $N-k$ drops out in round k at price P_k, with bidder 1 winning the auction at the final observed dropout price P_{N-2}.[8]

In ascending auctions, a Bayesian-Nash equilibrium consists of bid functions $\beta_i^k(X_i; \Omega_k)$ for each bidder i and round k, where Ω_k is the available information set at the beginning of round k containing the previously observed dropout prices. Effectively, the bidding function $\beta_i^k(X_i; \Omega_k)$ determines the price at which bidder i should quit the auction at round k as a function of his signal and the available information set. The collection of bid functions

[7] This standard model has been widley used by most of the subsequent literature (see Athey and Haile (2002), Hong and Shum (2003) and Aradillas-Lopez et al (2013) for examples).

[8] Given that continuous bidding does not take place in practice, I assign the dropout price of a bidder to the bid of the next bidder who outbids him.

$\beta_i^0(X_i;\Omega_0),\ldots,\beta_i^{N-2}(X_i;\Omega_{N-2})$ are common knowledge, with $\Omega_0=\emptyset$.

Like Milgrom and Weber (1982), I assume that the utility of bidder i depends only on the difference between his own valuation of the item put up for auction and his bid. More precisely, let $u[V_i-\beta_i^k(X_i;\Omega_k)]$ denote bidder i's utility at round k, where $u(\cdot)$ is continuous, nondecreasing in its argument and satisfies $u(0)=0$. But instead of an expected utility framework, as in the standard literature, I draw inspiration from the work in Kahneman and Tversky (1979) by assuming the following functional form:

$$u[V_i-\beta_i^k(X_i;\Omega_k)]=\begin{cases} V_i-\beta_i^k(X_i;\Omega_k) & \text{for} \quad V_i>\beta_i^k(X_i;\Omega_k) \\ \lambda_i[V_i-\beta_i^k(X_i;\Omega_k)] & \phantom{\text{for}} \quad V_i\leq\beta_i^k(X_i;\Omega_k) \end{cases}, \tag{1.1}$$

where $\lambda_i\geq 1$ captures loss aversion, i.e. the tendency of individuals to prefer avoiding reductions in wealth than equivalent gains. This piecewise linear specification, which has a kink at the origin,[9] has been used by many authors in a variety of economic situations (see Barberis et al (2001), Kőszegi and Rabin (2006) and Sprenger (2015) for examples). The reason is that loss aversion at the kink is very relevant for gambles that can lead to both gains and losses, such as in single item auctions, where "gains" and "losses" correspond to the difference between the value of the item auctioned and the final price.[10]

(Figure 1.1)

Figure 1.1 illustrates the effects of varying the loss aversion parameter λ on the underlying utility function (1.1). As expected, $\lambda=1$ implies risk neutrality, i.e. same marginal utility for both gains and losses (the standard model). However, for any other value of $\lambda>1$, bidders are more sensitive to reductions in wealth than to increases of the same magnitude, preferring not to lose \$10 rather than to gain \$10.

The structure of the Bayesian-Nash equilibrium of this asymmetric ascending auction in increasing bidding strategies (i.e. $\beta_i^k(X_i;\Omega_k)$ is increasing in X_i for $k=0,\ldots,N-2$) extends the equilibrium described in Milgrom and Weber (1982) and Hong and Shum (2003) to loss aversion as follows. For bidders $i=1,\ldots,N$ active in round 0, the bid functions are implicitly

[9] As in Kahneman and Tversky (1979), the primary reference level is the status quo, which in this case is 0, i.e. not participating in the auction.

[10] Kahneman and Tversky (1979) also propose that the utility function should be mildly concave over gains and convex over losses. However, this is most relevant when choosing between prospects that involve only gains or only losses (see Barberis et al (2001) for further discussion of this point).

defined by the equilibrium condition

$$E\{u[V_i - \beta_i^0(X_i;\Omega_0)]|\Upsilon_i^0\} = u\,(0) = 0,$$

where $\Upsilon_i^0 = \{X_i; X_j = \varphi_j^0[\beta_i^0(X_i;\Omega_0);\Omega_0]\}$ for $j = 1,\ldots,N$ and $j \neq i$, with $\varphi_j^k(.;\Omega_k)$ being the inverse bid function at round $k = 0,\ldots,N-2$ mapping prices into signals, so that $\varphi_i^k[\beta_i^k(X_i;\Omega_k);\Omega_k] = X_i$.

In turn, the analogous condition for bidders $i = 1,\ldots,N-k$ active in round $k = 1,\ldots,N-2$ will be given by

$$E\{u[V_i - \beta_i^k(X_i;\Omega_k)]|\Upsilon_i^k\} = u\,(0) = 0, \tag{1.2}$$

where $\Upsilon_i^k = \{X_i; X_j = \varphi_j^k[\beta_i^k(X_i;\Omega_k);\Omega_k], X_h = \varphi_h^{N-h}(P_{N-h};\Omega_{N-h})\}$, for $j = 1,\ldots,N-k$, $j \neq i$ and $h = N-k+1,\ldots N$, with X_h denoting the signals of the bidders who have dropped out prior to round k. Since the equilibrium bid functions are common knowledge, an active bidder in round k can infer the private information possessed by the previous dropout bidders by inverting their bid functions, so that $X_h = \varphi_h^{N-h}(P_{N-h};\Omega_{N-h})$.

Finally, it is worth mentioning that if several bidders were to quit simultaneously, the equilibrium conditions in (1.2) would still hold (see Milgrom and Weber (1982) for more details).

1.3.1 *The stochastic setup*

Following Hong and Shum (2003), I use a parametric approach by assuming that bidder's signals and valuations $(X_1,\ldots,X_N,V_1,\ldots,V_N)$ are log-normally distributed. This assumption allows me to derive tractable closed-form formulas for the expectations in (1.2), from which I can then obtain analytic expressions for the equilibrium bid functions $\beta_i^k(X_i;\Omega_k)$.[11]

Let V_i be defined as $V_i = A_i \times V$, where A_i is a bidder-specific private value component and V a common value component to all bidders in the auction. Although A_i and V, or indeed V_i, are not directly observed by the bidders, they are assumed to be independently log-normally distributed as follows:

$$\ln V = v = m + \epsilon_v \sim N(m, r_0^2),$$
$$\ln A_i = a_i = \bar{a}_i + \epsilon_{a_i} \sim N(\bar{a}_i, t_i^2),$$

[11] Two other empirical studies have previously used Hong and Shum's (2003) auction model. Dionne et al (2009) studied Mauritanian slave auctions in the 19th century, finding evidence of heterogeneity in the quality of the information between bidders, which in turn led to adverse selection. In turn, Koptyug (2016) found that in online car auctions, resellers are better than consumers at appraising the value of the cars they are bidding on.

so that

$$\ln V_i = v_i = \ln V + \ln A_i \sim N(m + \bar{a}_i, r_0^2 + t_i^2).$$

In practice, bidder i only observes a private noisy signal X_i of his own valuation V_i, which will be effectively revealed to the other bidders after he drops out. Given the log-normality assumption,

$$\ln X_i = x_i = v_i + \xi_i \sim N(m + \bar{a}_i, r_0^2 + t_i^2 + s_i^2),$$

where $\xi_i \sim N\left(0, s_i^2\right)$, and s_i^2 captures the amount of information any bidder has about the true value of the item being auctioned (see Dionne et al (2009) and Koptyug (2016) for more details). The common knowledge assumption implies that all the model parameters $\theta \equiv (\bar{a}_i, m, t_i^2, r_0^2, s_i^2)$ are known among the bidders.

In this log-normal setup, the conditional expected value of V_i can be written as:

$$E(V_i|X_1, \ldots, X_N) = \exp[E(v_i|x) + \frac{1}{2}Var(v_i|x)],$$

where $x = (x_1, \ldots, x_N)$,

$$E(v_i|x) \equiv \varsigma_{v_i|x} = v_i + \sigma'_{v_i x}\Sigma^{*-1}\left(x - \Psi\right)$$

and

$$Var(v_i|x) \equiv \omega_{v_i|x} = \sigma^2_{v_i} - \sigma'_{v_i x}\Sigma^{*-1}\sigma_{v_i x},$$

with

$$\mu_i = \begin{pmatrix} v_i & \Psi \end{pmatrix} \text{ and } \Sigma_i = \begin{pmatrix} \sigma^2_{v_i} & \sigma'_{v_i x} \\ \sigma_{v_i x} & \Sigma^* \end{pmatrix}$$

denoting the unconditional mean vector and variance-covariance matrix of $(v_i, x_1, x_2, \ldots, x_N)$ for bidder i (see section 1.7.2 for further details).

The following proposition, which I prove in section 1.7.1, establishes sufficient conditions to ensure the existence of an equilibrium under loss aversion in this stochastic framework.

Proposition 1.1 *Let $\eta_i \geq 0$ be the unique solution to*

$$\exp(\eta_i)\left[(1-\lambda_i)\,\mathrm{erf}\left(\frac{\frac{\omega_{v_i|x}}{2} + \eta_i}{\sqrt{2\omega_{v_i|x}}}\right) + (1+\lambda_i)\right] = \left[(\lambda_i - 1)\,\mathrm{erf}\left(\frac{\frac{\omega_{v_i|x}}{2} - \eta_i}{\sqrt{2\omega_{v_i|x}}}\right) + (\lambda_i + 1)\right],$$

where erf*(.) is the error function. Then*

$$\beta_i^k(X_i; \Omega_k) = \exp(-\eta_i)E(V_i|\Upsilon_i^k) \tag{1.3}$$

is an increasing-strategy Bayesian-Nash equilibrium under loss aversion in the log-normal stochastic setup.

(Figure 1.2)

Figure 1.2 compares the equilibrium bidding function under risk neutrality (the standard model) with loss aversion when $\lambda = 2.25$, a value based on the experimental findings in Tversky and Kahneman (1992). This graph shows that, ceteris paribus, loss aversion leads to a substantial reduction in the bids as a function of the signal X_i. As a consequence, the expected seller revenue will decrease relative to risk neutrality. However, the only difference between an equilibrium under risk neutrality and loss aversion is the multiplicative factor $\exp(-\eta_i)$ (see section 1.7.3 for more details).

To define the equilibrium log-bid functions for round k, I use the same notation as Hong and Shum (2003). Let $x_r^k = (x_1, \ldots, x_{N-k})'$ denote the vector of (log) private noisy signals of the bidders active in round k, and $x_d^k = (x_{N-k+1}, \ldots, x_N)'$ the vector of (log) signals of the dropped out bidders before round k. In addition, partition the inverse of the variance-covariance matrix of the private noisy signals as

$$\Sigma^{*-1} = \begin{pmatrix} \Sigma_{k,1}^{*-1} & \Sigma_{k,2}^{*-1} \end{pmatrix},$$

where $\Sigma_{k,1}^{*-1}$ is a $(N-k) \times N$ matrix corresponding to the remaining active bidders in round k, and $\Sigma_{k,2}^{*-1}$ is a $k \times N$ matrix corresponding to the bidders who have dropped out prior to round k.

Moreover, let

$$\begin{aligned} \Gamma_k &= \begin{pmatrix} \sigma_{\upsilon_1}^2 & \cdots & \sigma_{\upsilon_{N-k}}^2 \end{pmatrix}', \\ \Lambda_k &= \begin{pmatrix} \sigma_{\upsilon_1 x} & \cdots & \sigma_{\upsilon_{N-k} x} \end{pmatrix}', \\ \mu_k &= \begin{pmatrix} \upsilon_1 & \cdots & \upsilon_{N-k} \end{pmatrix}', \end{aligned}$$

and ℓ_k a $(N-k) \times 1$ vector of ones.

Additionally, let $\mathcal{A}^k$ and $\mathcal{C}^k$ be two $(N-k) \times 1$ vectors, and $\mathcal{D}^k$ a $(N-k) \times k$ matrix, with

$$\mathcal{A}^k = (\Lambda_k \Sigma_{k,1}^{*-1\prime})^{-1} \ell_k, \tag{1.4}$$

$$\mathcal{C}^k = \frac{1}{2}(\Lambda_k \Sigma_{k,1}^{*-1\prime})^{-1}[\Gamma_k - diag(\Lambda_k \Sigma^{*-1} \Lambda_k') + 2\mu_k - 2(\Lambda_k \Sigma^{*-1} \Psi)], \tag{1.5}$$

$$\mathcal{D}^k = (\Lambda_k \Sigma_{k,1}^{*-1\prime})^{-1}(\Lambda_k \Sigma_{k,2}^{*-1\prime}), \tag{1.6}$$

where $diag\,(.)$ is a matrix whose entries outside the main diagonal are all zero.

With this notation, the log-bidding function for the bidders active in round k under loss aversion will be

$$b_i^k(x_i; x_d^k) = \log[\beta_i^k(X_i; \Omega_k)] = \frac{1}{\mathcal{A}_i^k}(x_i + \mathcal{D}_i^k x_d^k + \mathcal{C}_i^k) - \eta_i,\ i = 1, \ldots, N-k, \tag{1.7}$$

where η_i captures the effects of loss aversion in (1.3), $\mathcal{A}_i^k$ and $\mathcal{C}_i^k$ denote the ith elements of the vectors (1.4) and (1.5), and $\mathcal{D}_i^k$ denotes the ith row of the matrix (1.6). Note that (1.7) depends on a bidder's own private signal x_i, as well as the signals of those bidders who have dropped out prior to round k x_d^k, except for round 0, where $\mathcal{D}^0$ and x_d^0 are obviously undefined.

Intuitively, by observing the dropout prices in previous rounds, the remaining active bidders can make inferences about the private information possessed by the bidders who have dropped out. In other words, they can obtain an unbiased estimate of bidder j's valuation from observing his private signal x_j. In common value auctions in which there is correlation across bidders' valuations (V_i), this information allows the remaining active bidders to update their beliefs about their own valuation, causing the prices at which bidders intend to exit to change as the auction progresses. In contrast, Vickrey (1961) showed that in private value auctions this updating does not occur, and each bidder has a weakly dominant strategy which is to bid up to his valuation (see Athey and Haile (2002) and section 1.7.4 for a more detailed discussion on private and common values auctions).

(Figure 1.3)

Figure 1.3 plots the equilibrium log-bid functions of a representative bidder in an auction with 5 loss averse bidders. The log-signal x_i is plotted on the x axis, while the log-bid functions in (1.7) for each round $k = 0, \ldots, 3$ are plotted on the y axis. As depicted in the figure, the slope of the log-bid function decreases for subsequent rounds, implying that, for a given realization of x_i, the targeted dropout price of the representative bidder decreases as the auction progresses. This occurs because bidders can update their bidding functions accordingly each round after incorporating the private noisy signal of the bidders who have previously drop out, thereby mitigating the chance of suffering the so-called winner's curse (see section 1.7.5 for more details).

1.3.2 *Econometric methodology*

Even though the model parameters $\theta \equiv (\bar{a}_i, m, t_i^2, r_0^2, s_i^2)$ are assumed to be known by the bidders, their values are unknown from an econometrician's point of view. In that regard, Hong and Shum (2003), Dionne et al (2009) and Koptyug (2016) employ the simulated non-linear least squares (SNLS) estimator of Laffont et al (1995), but with an independent probit kernel-smoother as in McFadden (1996). However, this estimation method does not always identify the parameters of the model for a small number of bidders. In contrast, Maximum likelihood (ML) identifies all the parameters even when there are only two bidders.

Furthermore, when the structural auction model is correctly specified, ML is more efficient than SNLS, but when it is misspecified, SNLS is not more robust than ML given that one must draw prices from the assumed model (see Dridi et al (2007) for more details). For example, suppose one estimates an independent private value model when in fact the true model is a pure common value one. In that case, both the log-bidding functions and the simulated drop out prices will be incorrect, which affects ML and SNLS.

Finally, Hong and Shum (2003) crucially show that the support of the (log) private signals (x) does not depend on θ in the log-normal stochastic setup in section 1.3.1, so the usual ML regularity conditions hold and standard asymptotic theory applies. For all these reasons, I will use ML to estimate the model.

In each auction, an econometrician only observes the vector of dropout prices for bidders $2, \ldots, N$, the order in which bidders drop out and their identities. As a consequence, Hong and Shum (2003) make clear that one must condition on the observed dropout sequence to derive the log-likelihood function. In practice, this means that the underlying log-signals $(x_1, \ldots, x_N)$ must be constrained to some region $\mathcal{T}_1(\theta) \subset R^N$, which I describe in section 1.7.6. Furthermore, they also show that as the winner's dropout bid is not observed, the winner's log-signal x_1 is constrained to some other region $\mathcal{T}_2(x_2, \ldots, x_N|\theta) \subset R^1$, which is consistent with bidder 1 winning the auction. Therefore, if $\mathcal{P} = (p_0, \ldots, p_{N-2})'$ denotes the vector of log-dropout bids, the log-likelihood function for a given auction must be computed as:

$$\mathcal{L}(\mathcal{P}|\theta) = \log f(\mathcal{P}|\theta) + \log \Pr[\mathcal{T}_2(x_2, \ldots, x_N|\theta)] - \log \Pr[\mathcal{T}_1(\theta)|\theta], \tag{1.8}$$

which resembles the log-likelihood function of a truncated and censored multivariate normal, with $f(\mathcal{P}|\theta)$ reflecting the continuous component corresponding to the likelihood of the observed drop out prices, $\Pr[\mathcal{T}_2(x_2,\ldots,x_N|\theta)]$ the conditional probability associated to the censored winning bid, and $\Pr[\mathcal{T}_1(\theta)|\theta]$ the truncation probability that reflects the order in which the different bidders drop out (see Hong and Shum (2003) and section 1.7.6 for more details on $\Pr[\mathcal{T}_1(\theta)|\theta]$, $\Pr[\mathcal{T}_2(x_2,\ldots,x_N|\theta)]$ and $f(\mathcal{P}|\theta)$).

Since the auctions take place independently, the sample log-likelihood function is the sum of the log-likelihood function of each auction. Thus, it is straightforward to combine auctions with different number of bidders.

From a practical point of view, the main difficulty in computing the log-likelihood function (1.8) is the multivariate integral $\Pr[\mathcal{T}_1(\theta)|\theta]$ (see again section 1.7.6 for details). Nevertheless, this is certainly feasible with up to 7 active bidders, although it slows down the numerical optimization. Still, given that the likelihood function is highly non-linear, it is convenient to consider multiple initial values.

1.4 The Information of Active Non-Bidding Participants

1.4.1 *The model*

A standard assumption in auction theory is that the bidders present at the auction coincide with all the potential bidders willing to participate (see Paarsch (1997), Krasnokutskaya and Seim (2011), Athey et al (2011) and Gentry and Li (2014) for examples).[12] Nevertheless, not all the bidders who are present in an ascending auction end up participating after observing the item put up for auction. In fact, some bidders decide not to participate when the auctioneer announces the opening bid. Therefore, it is important to distinguish between active bidders and active non-bidding participants in the following sense: active bidders are the ones who bid in the auction and either win or dropout at some point; in contrast, active non-bidding participants are the ones who are present in the auction but effectively drop at the opening bid.

All potential bidders observe each other when they assess the valuation of the item before the auction starts. Therefore, it seems reasonable to assume that at the beginning of the auction, active bidders can recover the private information active non-bidding participants possess and update their bidding functions accordingly.[13]

To define the equilibrium log-bid functions in this more general framework, let q denote the number of active non-bidding participants and N the number of active bidders, with $N+q$ being

[12] This assumption is not plausible in eBay auctions, as shown in Song (2004).

[13] In fact, the main bidders in *Storage Wars* usually publicly indicate their willigness to participate after oberving the locker to be auctioned.

the total number of potential bidders. At round -1, i.e. before the auction starts, let

$$\begin{array}{rcl} \dot{\Gamma}_{-1} & = & (\begin{array}{ccc} \sigma^2_{v_1} & \cdots & \sigma^2_{v_{N+q}} \end{array})', \\ \dot{\Lambda}_{-1} & = & (\begin{array}{ccc} \sigma_{v_1 x} & \cdots & \sigma_{v_{N+q} x} \end{array})', \\ \dot{\mu}_{-1} & = & (\begin{array}{ccc} v_1 & \cdots & v_{N+q} \end{array})', \end{array}$$

and $\dot{\ell}_{-1}$ a $(N+q) \times 1$ vector of ones.

Additionally, let $\dot{\mathcal{A}}^{-1}$ and $\dot{\mathcal{C}}^{-1}$ be two $(N+q) \times 1$ vectors, with

$$\begin{array}{rcl} \dot{\mathcal{A}}^{-1} & = & (\dot{\Lambda}_{-1}\Sigma^{*-1\prime})^{-1}\dot{\ell}_{-1}, \\ \dot{\mathcal{C}}^{-1} & = & \frac{1}{2}(\dot{\Lambda}_{-1}\Sigma^{*-1\prime})^{-1}[\dot{\Gamma}_{-1} - diag(\dot{\Lambda}_{-1}\Sigma^{*-1}\dot{\Lambda}'_{-1}) + 2\dot{\mu}_{-1} - 2(\dot{\Lambda}_{-1}\Sigma^{*-1}\Psi)]. \end{array}$$

With this notation, the log-bidding function for the bidders in round -1 is given by:

$$b_i^{-1}(x_i) = \log[\beta_i^{-1}(X_i;\dot{\Omega}_{-1})] = \frac{1}{\dot{\mathcal{A}}_i^{-1}}(x_i + \dot{\mathcal{C}}_i^{-1}) - \eta_i, \; i = 1, \ldots, N+q,$$

which again reflects loss aversion, captured by η_i in expression (1.3), and depends only on bidder's i own private signal x_i. In fact, $b_i^{-1}(x_i)$ is equivalent to the log-bidding functions in round 0 without active non-bidding participants in (1.7).

In this context, the active non-bidding participants will be the ones who on the basis of this log-bidding function decide not to participate when the auctioneer announces the opening bid P_{-1}.

Since the equilibrium bid functions are common knowledge, at round 0 active bidders can use the information the q active non-bidding participants possessed to infer their own private signals by inverting the log-bid functions of the active non-bidding participants. Thus,

$$x_q = \ln X_q = b_q^{-1}(x_q)\dot{\mathcal{A}}_q^{-1} - \dot{\mathcal{C}}_q^{-1}.$$

Let $\ddot{x}_r^0 = (x_1, \ldots, x_N)'$ denote the vector of private noisy signals of the active bidders in round 0, and $\ddot{x}_d^0 = (x_{N+1}, \ldots, x_{N+q})'$ the vector of signals of the active non-bidding participants who effectively dropped out in round -1. In addition, partition the inverse of the variance-covariance matrix of the private noisy signals as

$$\Sigma^{*-1} = (\begin{array}{cc} \ddot{\Sigma}_{0,1}^{*-1} & \ddot{\Sigma}_{0,2}^{*-1} \end{array}),$$

where $\ddot{\Sigma}_{0,1}^{*-1}$ is a $N \times (N+q)$ matrix corresponding to the active bidders in round 0, and $\ddot{\Sigma}_{0,2}^{*-1}$

is a $q \times (N+q)$ matrix corresponding to the active non-bidding participants who dropped out in round -1.

Moreover, let

$$\begin{aligned}
\ddot{\Gamma}_0 &= \begin{pmatrix} \sigma^2_{v_1} & \cdots & \sigma^2_{v_N} \end{pmatrix}', \\
\ddot{\Lambda}_0 &= \begin{pmatrix} \sigma_{v_1 x} & \cdots & \sigma_{v_N x} \end{pmatrix}', \\
\ddot{\mu}_0 &= \begin{pmatrix} v_1 & \cdots & v_N \end{pmatrix}',
\end{aligned}$$

and $\ddot{\ell}_0$ a $N \times 1$ vector of ones.

Additionally, let $\ddot{\mathcal{A}}^0$ and $\ddot{\mathcal{C}}^0$ be two $N \times 1$ vectors, and $\ddot{\mathcal{D}}^0$ a $N \times k$ matrix, with

$$\begin{aligned}
\ddot{\mathcal{A}}^0 &= (\ddot{\Lambda}_0 \ddot{\Sigma}^{*-1\prime}_{0,1})^{-1} \ddot{\ell}_0, \\
\ddot{\mathcal{C}}^0 &= \frac{1}{2} (\ddot{\Lambda}_0 \ddot{\Sigma}^{*-1\prime}_{0,1})^{-1} [\ddot{\Gamma}_0 - diag(\ddot{\Lambda}_0 \Sigma^{*-1} \ddot{\Lambda}_0') + 2\ddot{\mu}_0 - 2(\ddot{\Lambda}_0 \Sigma^{*-1} \Psi)], \\
\ddot{\mathcal{D}}^0 &= (\ddot{\Lambda}_0 \ddot{\Sigma}^{*-1\prime}_{0,1})^{-1} (\ddot{\Lambda}_0 \ddot{\Sigma}^{*-1\prime}_{0,2}).
\end{aligned}$$

With this notation, the log-bidding function for the active bidders in round 0 under loss aversion is:

$$b^0_i(x_i; \ddot{x}^0_d) = \log[\beta^0_i(X_i; \ddot{\Omega}_0)] = \frac{1}{\ddot{\mathcal{A}}^0_i} (x_i + \ddot{\mathcal{D}}^0_i \ddot{x}^0_d + \ddot{\mathcal{C}}^0_i) - \eta_i, \; i = 1, \ldots, N.$$

Compared to equation (1.7) for $k = 0$, which only depends on a bidder's own signal x_i, now $b^0_i(x_i; \ddot{x}^0_d)$ is also a function of the signals of the active non-bidding participants $\ddot{x}^0_d$.

For any subsequent round $k = 1, \ldots, N-2$, the log-bidding function for the bidders active in round k are entirely analogous to (1.7), and therefore depends on a bidder's own private signal x_i, as well as the signals of those bidders who have dropped out prior to round k, including the active non-bidding participants, i.e.

$$x^k_d = (\underbrace{x_{N-k+1}, \ldots, x_N}_{N-k}, \underbrace{x_{N+1}, \ldots, x_{N+q}}_{q})'.$$

(Figure 1.4)

Figure 1.4 compares the equilibrium log-bidding function for a loss averse active bidder in round 0 with 0, 1 and 2 active non-bidding participants. As expected, bidders reduce the aggressiveness of their bids even further as the number of active non-bidding participants increases, substantially reducing the chances of suffering the winner's curse. Intuitively, this occurs because at round 0 active bidders recover the private information active non-bidding participants possess, and they update their log-bidding functions accordingly.[14]

1.4.2 *Econometric methodology*

The structure of the log-likelihood function is similar to the one in section 1.3.2. Therefore, conditional on the vector of active non-bidding participants' dropout bids, the log-likelihood function for a given auction can be written as:

$$\mathcal{L}(\mathcal{P}|\theta, \dot{\Upsilon}^{-1}) = \log f(\mathcal{P}|\theta, \dot{\Upsilon}^{-1}) + \log \Pr[\mathcal{T}_2(x_d^{N-2}|\theta, \dot{\Upsilon}^{-1})] - \log \Pr[\mathcal{T}_1(\theta)|\theta, \dot{\Upsilon}^{-1}],$$

where $\dot{\Upsilon}^{-1} = [X_l = \varphi_l^{-1}(P_{-1}; \Omega_{-1})]$ for $l = N+1, \ldots, N+q$, with $\varphi_l^{-1}(.; \Omega_{-1})$ being the vector of inverse bid functions at round -1, X_l denoting the signals of the q active non-bidding participants, $f(\mathcal{P}|\theta, \dot{\Upsilon}^{-1})$ reflecting the (conditional) continuous likelihood of the observed drop out prices, $\Pr[\mathcal{T}_2(x_d^{N-2}|\theta, \dot{\Upsilon}^{-1})]$ the conditional probability associated to the censored winning bid, and $\Pr[\mathcal{T}_1(\theta)|\theta, \dot{\Upsilon}^{-1}]$ the (conditional) truncation probability that reflects the order in which the different bidders drop out (see section 1.7.7 for more details on $f(\mathcal{P}|\theta, \dot{\Upsilon}^{-1})$, $\Pr[\mathcal{T}_2(x_d^{N-2}|\theta, \dot{\Upsilon}^{-1})]$ and $\Pr[\mathcal{T}_1(\theta)|\theta, \dot{\Upsilon}^{-1}]$).

1.5 Empirical Application

Figure 1.5 displays the boxplot of the profit/losses in *Storage Wars* auctions without a few extreme outliers.

(Figure 1.5)

The central mark in the box indicates the median profit (\$890), and the bottom and top edges indicate the 25th (\$-47.5) and 75th (\$2,412.5) percentiles, respectively, with the outliers being plotted using the + symbol. As can be seen, the profit/losses values involved in these auctions are relatively small. Therefore, the smooth utility functions with moderate risk aversion commonly considered in the literature under expected utility imply that bidders would be close to risk neutral when facing such modest stakes. In contrast, loss aversion may be present in these auctions because the utility in (1.1) captures the well documented fact that over modest gambles, individuals are noticeably more averse to losses relative to the status quo than they

[14] With homogeneous bidders, the log-bidding functions in Figure 1.4 are equivalent to the log-bidding functions without active non-bidding participants in Figure 1.3 for rounds 0, 1 and 2.

are attracted by gains (see Barberis et al (2001) for more details).

1.5.1 *Model specification*

Given that the common public information bidders have during the auction are the locker characteristics and the municipality in which they are located, I have regressed the (log) ex-post value of the locker on its size and the per capita income of the municipality. The results are presented in Table 1.4.[15]

(Table 1.4)

Not surprisingly, the statistical significance of the results confirm that richer neighborhoods and larger lockers have more valuable locker contents. Consequently, I specify the mean of the common value component for a given auction as

$$m = \beta_0 + \beta_1 SIZE + \beta_2 HHI,$$

where $SIZE$ is a variable that measures the size of the locker (small (1), medium (2) or large (3)) and HHI captures the median household income of the municipality where the locker is located in the State of California.

In contrast, private valuations are usually associated with differences in interests across bidders, for which I do not observe any proxies. For that reason, I flexibly define the mean of the private value component of the four main bidders (Barry "Ba", Darrell "Dr", Dave "Dv" and Jarrod "Jr"), as well as of the other active bidders whose identity is not shown publicly, as

$$\bar{a} = \begin{pmatrix} \alpha_0 + \alpha_1 Ba & \alpha_0 + \alpha_2 Dr & \alpha_0 + \alpha_3 Dv & \alpha_0 + \alpha_4 Jr & \alpha_0 & \cdots & \alpha_0 \end{pmatrix},$$

where Ba, Dr, Dv and Jr are mutually exclusive dummy variables. For example, Ba takes the value 1 if Barry is an active bidder in the auction and 0 otherwise. Note that α_0 is the common mean of the private value component of those active bidders whose identity is unknown.[16]

Furthermore, to guarantee positivity, the variance of the common value component, which is obviously the same across bidders, is modelled as $r_0^2 = \exp(\delta_0)$, while the variance of the noise

[15] A more flexible non-linear specification that allows for different coefficients for each of the three locker sizes does not offer any statistically significant gains in fit, which is not surprinsing given that the sequence of locker sizes corresponds to 3, 5 and 7 rooms (see section 1.2.2 for more details).

[16] Given that in all the formulas all that matters is $m + \bar{a}_i$ (see section 1.7.2 for further details), I set $\beta_0 = 0$ without loss of generality because the constant terms of $\bar{a}$ and m are not separately identified.

for each of the bidder's signals is flexibly defined as

$$s^2 = \exp(\; \gamma_0 + \gamma_1 Ba \quad \gamma_0 + \gamma_2 Dr \quad \gamma_0 + \gamma_3 Dv \quad \gamma_0 + \gamma_4 Jr \quad \gamma_0 \quad \cdots \quad \gamma_0 \;),$$

so that γ_0 is the baseline variance of the anonymous bidders.

In principle, I also allow for unrestricted heterogeneity in the variance of the private value component as follows

$$t^2 = \exp(\; \tau_0 + \tau_1 Ba \quad \tau_0 + \tau_2 Dr \quad \tau_0 + \tau_3 Dv \quad \tau_0 + \tau_4 Jr \quad \tau_0 \quad \cdots \quad \tau_0 \;).$$

Finally, I set the loss aversion parameter λ to 2.25, a value initially proposed by Tversky and Kahneman (1992) on the basis of experimental evidence which has been used by most of the subsequent literature (see for example Barberis et al (2001), Barberis and Huang (2008) and Post et al (2008)).

1.5.2 *Parameter estimates of the baseline model*

The first thing I do is check whether *Storage Wars* bidders exhibit loss aversion. To do so, I fit the model with $\lambda = 2.25$ for all the bidders and compare it to a specification with risk neutrality ($\lambda = 1$). Surprisingly, the likelihood is actually worse. However, given that the model in section 1.3 explicitly allows for heterogeneous bidders' characteristics, this two extreme specifications are not the only ones that one could consider. In fact, when I set $\lambda = 1$ for Dave and $\lambda = 2.25$ for all the other bidders, I find that the difference between the log-likelihoods of the risk neutral model and this alternative specification is 9.41, thus confirming the empirical relevance of loss aversion in ascending auctions.

As I explained in section 1.2.1, Dave is the most professional bidder in *Storage Wars*. Therefore, my finding is not entirely surprising in view of the results in List (2004), who found that professional traders did not exhibit loss aversion.[17] In this respect, it is worth mentioning that Dave suffers the smallest median loss when he losses and enjoys the largest median profit when he wins, regardless of the size of the locker.

The maximum likelihood estimates of the model parameters for this specification are shown in Table 1.5.

(Table 1.5)

[17] In contrast, Pope and Schweitzer (2011) found that even the best golfers seem loss averse in the non-pecuniary context of golf putts.

The results indicate that the coefficients of the size of the locker (β_1) and the per capita income of the municipality (β_2) are both positive and statistically significant, which agrees with the findings in Table 1.4 regarding the specification of the mean of the common value component (m). Additionally, there is strong evidence of asymmetry in terms of the mean of the private value components (p-value of 0 for LR test of $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$) and weaker evidence of heterogeneity in the accuracy of the signals (p-value of 0.07 for LR test of $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0$). However, there is no evidence of heterogeneity in the importance of the private value component when t^2 is heterogeneously modelled as in section 1.5.1 (p-value of 0.43 for LR test of $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$).

The variance of the common value component (r_0^2) explains 74% of the variance of the valuation V_i (see section 1.3.1 and section 1.7.2), which reflects that the model is neither a pure common value nor an independent private value model, but a mixture of both. To confirm this claim, I formally compare my estimated model to those two extreme versions:

	LR Test	p-value
Independent Private Value	404.23	0
Pure Common Value	77.24	0

Although the pure common value model provides a better match of the results in Table 1.5, it is still rejected by a long margin.

(Figure 1.6)

Figure 1.6 plots the estimated equilibrium log-bidding functions at round 0 of *Storage Wars* bidders, all of whom are loss averse except for Dave, who is risk neutral. As can be seen, Dave, whose marginal utility is the same for both gains and losses, is the most aggressive bidder for most signal values, although his bidding function has the lowest slope. At the opposite extreme, Barry is the least aggressive bidder. As an illustration, suppose both of them had the same log-signal $x_i = 8.5$ (\$4,914), which is approximately the average ex-post value of all the lockers in *Storage Wars*. Then, we can read off the graph that Dave's targeted log-dropout price in round 0 would be 8.48 (\$4,821), while it would be 6.91 (\$1,007) for Barry.

1.5.3 *Parameter estimates with active non-bidding participants*

Following the evidence in the previous section, I continue to set $\lambda = 1$ for Dave (the most professional bidder in the sample) and $\lambda = 2.25$ for all the other bidders. In this case, the improvement in the log-likelihood function relative to the risk neutral model is 12.73, which is even greater than in section 1.5.2. Therefore, loss aversion is again empirically relevant in this more general framework.

The maximum likelihood estimates of the model parameters for this specification are shown in Table 1.6.

(Table 1.6)

As in section 1.5.2, the values of β_1 and β_2 are statistically significant. Additionally, I find that there is strong evidence of heterogeneity in the precision of the signals (p-value of 0 for LR test of $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0$) and of asymmetry in the mean of the private value component (p-value of 0 for LR test of $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$). However, once again I find no evidence of heterogeneity in the importance of the private value component t^2 (p-value of 0.77 for LR test of $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$), as in Table 1.5.

(Figure 1.7)

Figure 1.7 illustrates the estimated equilibrium log-bidding functions of *Storage Wars* bidders in round 0 with 1 active non-bidding participant, in this case Darrell. This graph shows that the slope of the log-bidding functions for the remaining bidders decreases substantially compared to their round 0 log-bidding functions in Figure 1.6. Intuitively, this reflects the fact that they effectively take into account the private information of the bidder who decided not to participate, thereby confirming the empirical relevance of active non-bidding participants in ascending auctions.

1.6 Conclusions

In this chapter I propose a novel tractable structural model with both private and common value components for ascending auctions in which heterogeneous bidders may exhibit loss aversion. Importantly, I find that, ceteris paribus, the bidding functions of a loss averse bidder are significantly lower than under risk neutrality.

Additionally, I consider a more general framework in which active bidders incorporate into their strategies the information of those bidders who are present but decide not to participate after observing the item put up for auction when the auctioneer announces the opening bid. In this respect, I find that bidders reduce the aggressiveness of their bids even further as the number of non-bidding participants increases.

To empirically asses my model, I use data from the popular TV show *Storage Wars*, which follows some recurrent individual bidders who take part in storage locker auctions throughout the State of California.

My empirical results document for the first time the presence of loss aversion in actual ascending auctions. More precisely, I find that the behavior of most bidders is consistent with loss aversion in a model in which there is heterogeneity in both the mean of the private value component and the precision of the signals. At the same time, I find that the most professional bidder seems to be risk neutral.

I also find that loss aversion persists when bidders incorporate into their strategies the information of those bidders who are present but decide not to participate after observing the item to be auctioned. Moreover, my findings confirm the empirical relevance of taking into account the presence of non-bidding participants in ascending auctions.

Although the empirical analysis of this chapter provides reliable evidence of the importance of loss aversion in ascending auctions, there is still much to learn about the behavioral biases that arise in auctions from the field, lab and real life situations.

1.7 Proofs and Auxiliary Results

1.7.1 *Equilibrium proof*

For notational simplicity, I suppress the arguments of the bid functions so that $\beta_i^k(X_i;\Omega_k) = \beta_i^k(\cdot)$.

Following the discussion in (1.2), for any round k,

$$E\{u[V_i - \beta_i^k(\cdot)]|\Upsilon_i^k\} = \int_0^{\beta_i^k(\cdot)} \{\lambda_i[V_i - \beta_i^k(\cdot)] \times f[V_i - \beta_i^k(\cdot)|\Upsilon_i^k]\}d[V_i - \beta_i^k(\cdot)] \\ + \int_{\beta_i^k(\cdot)}^{+\infty} \{[V_i - \beta_i^k(\cdot)] \times f[V_i - \beta_i^k(\cdot)|\Upsilon_i^k]\}d[V_i - \beta_i^k(\cdot)].$$

Since $f[V_i - \beta_i^k(\cdot)|\Upsilon_i^k] = f(V_i|\Upsilon_i^k)$, then $E\{u[V_i - \beta_i^k(\cdot)]|\Upsilon_i^k\}$ can be written as

$$E\{u[V_i - \beta_i^k(\cdot)]|\Upsilon_i^k\} = \int_0^{\beta_i^k(\cdot)} \{\lambda_i[V_i - \beta_i^k(\cdot)] \times f(V_i|\Upsilon_i^k)\}dV_i \\ + \int_{\beta_i^k(\cdot)}^{+\infty} \{[V_i - \beta_i^k(\cdot)] \times f(V_i|\Upsilon_i^k)\}dV_i.$$

Given that $V_i = \exp(v_i)$, where $v_i \sim N[E(v_i), Var(v_i)]$, the density of V_i is

$$f(V_i|\Upsilon_i^k) = \frac{1}{V_i\sqrt{2\pi Var(v_i)}} \exp\left[-\frac{[\log(V_i) - E(v_i)]^2}{2Var(v_i)}\right].$$

Moreover,

$$\int_0^{\beta_i^k(\cdot)} \{\lambda_i[V_i - \beta_i^k(\cdot)] \times f(V_i|\Upsilon_i^k)\}dV_i = \int_0^{\beta_i^k(\cdot)} \frac{\Pr[0 < V_i < \beta_i^k(\cdot)]}{\Pr[0 < V_i < \beta_i^k(\cdot)]}\{\lambda_i[V_i - \beta_i^k(\cdot)] \times f(V_i|\Upsilon_i^k)\}dV_i$$

or equivalently

$$\int_0^{\beta_i^k(\cdot)} \{\lambda_i[V_i - \beta_i^k(\cdot)] \times f(V_i|\Upsilon_i^k)\}dV_i = \{\lambda_i[\Pr 0 < V_i < \beta_i^k(\cdot)\} \\ \times \left[\int_0^{\beta_i^k(\cdot)} \frac{[V_i \times f(V_i|\Upsilon_i^k)]}{\Pr[0 < V_i < \beta_i^k(\cdot)]}dV_i - \beta_i^k(\cdot)\int_0^{\beta_i^k(\cdot)} \frac{f(V_i|\Upsilon_i^k)}{\Pr[0 < V_i < \beta_i^k(\cdot)]}dV_i\right],$$

with $\int_0^{\beta_i^k(\cdot)}\{f(V_i|\Upsilon_i^k)/\Pr[0 < V_i < \beta_i^k(\cdot)]\}dV_i = 1$.

Note that $\Pr[0 < V_i < \beta_i^k(\cdot)] = \Pr[\ln(0) < v_i < \ln(\beta_i^k(\cdot)]$, so

$$\Pr[0 < V_i < \beta_i^k(\cdot)] = \Phi\left(\frac{\ln[\beta_i^k(\cdot)] - \varsigma_{v_i|x}}{\sqrt{\omega_{v_i|x}}}\right) = \frac{1}{2}\left(\text{erf}\left\{\frac{\ln[\beta_i^k(\cdot)] - \varsigma_{v_i|x}}{\sqrt{2\omega_{v_i|x}}}\right\} + 1\right)$$

Therefore,

$$\int_0^{\beta_i^k(\cdot)} \{\lambda_i[V_i - \beta_i^k(\cdot)] \times f(V_i|\Upsilon_i^k)\}dV_i = \{\lambda_i \Pr[0 < V_i < \beta_i^k(\cdot)]\}\left\{\int_0^{\beta_i^k(\cdot)} \frac{[V_i \times f(V_i|\Upsilon_i^k)]}{\Pr[0 < V_i < \beta_i^k(\cdot)]}dV_i\right\}.$$

Following Zaninetti (2017),

$$\int_0^{\beta_i^k(\cdot)} \frac{\left[V_i \times f(V_i|\Upsilon_i^k)\right]}{\Pr[0 < V_i < \beta_i^k(\cdot)]} dV_i = E_{T1},$$

where

$$E_{T1} = \frac{\exp[\frac{1}{2}Var(v_i)]\exp[E(v_i)]\,[\mathrm{erf}(a_1) + \mathrm{erf}(a_2)]}{[\mathrm{erf}(a_3) + \mathrm{erf}(a_4)]},$$

with $a_1 = [-\infty - \omega_{v_i|x} - \varsigma_{v_i|x}]/\sqrt{2\omega_{v_i|x}}$, $a_2 = \{\omega_{v_i|x} + \varsigma_{v_i|x} - \ln[\beta_i^k(\cdot)]\}/\sqrt{2\omega_{v_i|x}}$, $a_3 = [-\infty - \varsigma_{v_i|x}]/\sqrt{2\omega_{v_i|x}}$ and $a_4 = \{\varsigma_{v_i|x} - \ln[\beta_i^k(\cdot)]\}/\sqrt{2\omega_{v_i|x}}$.

Hence,

$$E_{T1} = \frac{\exp\left(\frac{1}{2}\omega_{v_i|x}\right)\exp(\varsigma_{v_i|x})\,[\mathrm{erf}(a_2) - 1]}{[\mathrm{erf}(a_4) - 1]},$$

because $\mathrm{erf}(a_1) = \mathrm{erf}(a_3) = -1$.

Therefore,

$$\int_0^{\beta_i^k(\cdot)} \{\lambda_i[V_i - \beta_i^k(\cdot)] \times f(V_i|\Upsilon_i^k)\} dV_i = \{\lambda_i[\Pr 0 < V_i < \beta_i^k(\cdot)]\}[E_{T1} - \beta_i^k(\cdot)].$$

Similarly,

$$\begin{aligned}\int_{\beta_i^k(\cdot)}^{+\infty} \{[V_i - \beta_i^k(\cdot)] \times f(V_i|\Upsilon_i^k)\} dV_i &= \Pr[\beta_i^k(\cdot) < V_i < +\infty] \\ &\times \left\{ \int_{\beta_i^k(\cdot)}^{+\infty} \frac{\left[V_i \times f(V_i|\Upsilon_i^k)\right]}{\Pr[0 < V_i < \beta_i^k(\cdot)]} dV_i - \beta_i^k(\cdot) \int_0^{\beta_i^k(\cdot)} \frac{f(V_i|\Upsilon_i^k)}{\Pr[0 < V_i < \beta_i^k(\cdot)]} dV_i \right\}.\end{aligned}$$

But since $\Pr[\beta_i^k(\cdot) < V_i < +\infty] = \Pr\{\ln[\beta_i^k(\cdot)] < v_i < +\infty\}$, then

$$\Pr[\beta_i^k(\cdot) < V_i < +\infty] = \frac{1}{2}\left(1 - \left\{\mathrm{erf}\,\frac{\ln[\beta_i^k(\cdot)] - \varsigma_{v_i|x}}{\sqrt{2\omega_{v_i|x}}}\right\}\right).$$

Hence,

$$\int_{\beta_i^k(\cdot)}^{+\infty} \{[V_i - \beta_i^k(\cdot)] \times f(V_i|\Upsilon_i^k)\} dV_i = \Pr[\beta_i^k(\cdot) < V_i < +\infty] \left[\int_{\beta_i^k(\cdot)}^{+\infty} \frac{\left[V_i \times f(V_i|\Upsilon_i^k)\right]}{\Pr[0 < V_i < \beta_i^k(\cdot)]} dV_i \right].$$

Again, following Zaninetti (2017),

$$\int_{\beta_i^k(\cdot)}^{+\infty} \frac{\left[V_i \times f(V_i|\Upsilon_i^k)\right]}{\Pr[0 < V_i < \beta_i^k(\cdot)]} dV_i = E_{T2},$$

where

$$E_{T2} = \frac{\exp[\frac{1}{2}Var(v_i)]\exp[E(v_i)]\,[\mathrm{erf}(b_1) + \mathrm{erf}(b_2)]}{[\mathrm{erf}(b_3) + \mathrm{erf}(b_4)]},$$

with $b_1 = -a_2$, $b_2 = [\omega_{v_i|x} + \varsigma_{v_i|x} - \infty]/\sqrt{2\omega_{v_i|x}}$, $b_3 = -a_4$ and $b_4 = [\varsigma_{v_i|x} - \infty]/\sqrt{2\omega_{v_i|x}}$.

Hence,

$$E_{T2} = \frac{\exp\left(\frac{1}{2}\omega_{v_i|x}\right)\exp(\varsigma_{v_i|x})\left[-\operatorname{erf}(a_2) - 1\right]}{\left[-\operatorname{erf}(a_4) - 1\right]},$$

because $\operatorname{erf}(b_2) = \operatorname{erf}(b_4) = -1$.

Therefore,

$$\int_{\beta_i^k(\cdot)}^{+\infty} \{[V_i - \beta_i^k(\cdot)] \times f(V_i|\Upsilon_i^k)\} dV_i = \Pr[\beta_i^k(\cdot) < V_i < +\infty][E_{T2} - \beta_i^k(\cdot)],$$

so $E\{u[V_i - \beta_i^k(\cdot)]|\Upsilon_i^k\}$ is then

$$\begin{aligned} E\{u[V_i - \beta_i^k(\cdot)]|\Upsilon_i^k\} = \{\lambda_i \Pr[0 < V_i < \beta_i^k(\cdot)]\}[E_{T1} - \beta_i^k(\cdot)] \\ + \{\Pr[\beta_i^k(\cdot) < V_i < +\infty\}[E_{T2} - \beta_i^k(\cdot)]. \end{aligned}$$

In equilibrium,

$$\{\lambda_i \Pr[0 < V_i < \beta_i^k(\cdot)]\}[E_{T1} - \beta_i^k(\cdot)] + \{\Pr[\beta_i^k(\cdot) < V_i < +\infty]\}[E_{T2} - \beta_i^k(\cdot)] = 0,$$

which simplifies to

$$\begin{aligned} \exp\left(\frac{1}{2}\omega_{v_i|x}\right)\exp(\varsigma_{v_i|x})\left[(1-\lambda_i)\left\{\operatorname{erf}\frac{\omega_{v_i|x} + \varsigma_{v_i|x} - \ln[\beta_i^k(\cdot)]}{\sqrt{2\omega_{v_i|x}}}\right\} + (1+\lambda_i)\right] \\ - \exp\{\ln[\beta_i^k(\cdot)]\}\left[(\lambda_i - 1)\operatorname{erf}\left[\frac{\ln(P_k) - \varsigma_{v_i|x}}{\sqrt{2\omega_{v_i|x}}}\right] + (\lambda_i + 1)\right] = 0. \end{aligned}$$

When $\lambda_i > 1$ and $\lambda_i \neq 1$, by "Guess and Verify", it is clear that the solution is:

$$\varsigma_{v_i|x} = \ln[\beta_i^k(\cdot)] - \frac{1}{2}\omega_{v_i|x} + \eta_i,$$

with η_i solving

$$\exp(\eta_i)\left[(1-\lambda_i)\operatorname{erf}\left(\frac{\frac{\omega_{v_i|x}}{2} + \eta_i}{\sqrt{2\omega_{v_i|x}}}\right) + (1+\lambda_i)\right] - \left[(\lambda_i - 1)\operatorname{erf}\left(\frac{\frac{\omega_{v_i|x}}{2} - \eta_i}{\sqrt{2\omega_{v_i|x}}}\right) + (\lambda_i + 1)\right] = 0. \tag{1.9}$$

If there exists an η_i that solves (1.9), then the above solution solves the original system. To confirm this claim, let

$$\begin{aligned} Y(\eta_i) \equiv \exp(\eta_i)\left[(1-\lambda_i)\operatorname{erf}\left(\frac{\frac{\omega_{v_i|x}}{2} + \eta_i}{\sqrt{2\omega_{v_i|x}}}\right) + (1+\lambda_i)\right] \\ - \left[(\lambda_i - 1)\operatorname{erf}\left(\frac{\frac{\omega_{v_i|x}}{2} - \eta_i}{\sqrt{2\omega_{v_i|x}}}\right) + (\lambda_i + 1)\right] = 0. \end{aligned}$$

To check whether η_i is a solution to $Y(\eta_i) = 0$, one can exploit the fact that

$$\left.\begin{array}{l} 1)\ \lim_{\eta_i \to -\infty} Y(\eta_i) < 0 \\ 2)\ \lim_{\eta_i \to +\infty} Y(\eta_i) > 0 \end{array}\right\}$$

Specifically, given that $0 < \omega_{\upsilon_i|x} < \infty$ and $\lambda_i > 1$, then

$$\lim_{\eta_i \to -\infty} Y(\eta_i) = -2\lambda_i < 0 \quad \text{and} \quad \lim_{\eta_i \to +\infty} Y(\eta_i) = +\infty > 0.$$

As a special case,

$$\lim_{\eta_i \to 0} Y(\eta_i) = 2(1-\lambda_i)\,\mathrm{erf}\left(\frac{\frac{\omega_{\upsilon_i|x}}{2}}{\sqrt{2\omega_{\upsilon_i|x}}}\right) \le 0.$$

The continuity of $Y(\eta_i)$ guarantees that there exists an η_i that solves $Y(\eta_i) = 0$.

If in addition $\partial Y(\eta_i)/\partial\eta_i > 0$ for any $-\infty < \eta_i < \infty$, the solution will be unique. In particular,

$$\begin{aligned} \partial Y(\eta_i)/\partial\eta_i = \exp(\eta_i) & \left\{ (1-\lambda_i)\,\mathrm{erf}\left(\frac{\frac{\omega_{\upsilon_i|x}}{2}+\eta_i}{\sqrt{2\omega_{\upsilon_i|x}}}\right) + (1+\lambda_i) + \frac{2(1-\lambda_i)}{\sqrt{2\pi\omega_{\upsilon_i|x}}}\exp\left[-\left(\frac{\frac{\omega_{\upsilon_i|x}}{2}+\eta_i}{\sqrt{2\omega_{\upsilon_i|x}}}\right)^2\right]\right\} \\ & + \left\{\frac{2(\lambda_i-1)}{\sqrt{2\pi\omega_{\upsilon_i|x}}}\exp\left[-\left(\frac{\frac{\omega_{\upsilon_i|x}}{2}-\eta_i}{\sqrt{2\omega_{\upsilon_i|x}}}\right)^2\right]\right\} > 0, \end{aligned}$$

or equivalently

$$\frac{(1+\lambda_i)}{(\lambda_i-1)} - \mathrm{erf}\left(\frac{\frac{\omega_{\upsilon_i|x}}{2}+\eta_i}{\sqrt{2\omega_{\upsilon_i|x}}}\right) > 0,$$

which is true for any value of η_i and $0 < \omega_{\upsilon_i|x} < \infty$ because $[(1+\lambda_i)/(\lambda_i - 1)] > 1$ for $1 < \lambda_i < \infty$ and $\mathrm{erf}[(\frac{1}{2}\omega_{\upsilon_i|x} + \eta_i)/\sqrt{2\omega_{\upsilon_i|x}}] \in [-1, 1]$. As an aside, it is worth mentioning that η_i, despite being heterogeneous, does not depend on the round of the auction or on the bidder's own private signal.

Therefore, given the existence and uniqueness of η_i,

$$\varsigma_{\upsilon_i|x} = (\upsilon_i - \sigma'_{\upsilon_i x}\Sigma^{*-1}\Psi) + \sigma'_{\upsilon_i x}\Sigma^{*-1}_{k,1}x^k_r + \sigma'_{\upsilon_i x}\Sigma^{*-1}_{k,2}x^k_d,$$

where $\Sigma^{*-1} = (\Sigma^{*-1}_{k,1}, \Sigma^{*-1}_{k,2})$. Solving for x^k_r yields:

$$x^k_r = (\sigma'_{\upsilon_i x}\Sigma^{*-1}_{k,1})^{-1}[\varsigma_{\upsilon_i|x} + (\sigma'_{\upsilon_i x}\Sigma^{*-1}\Psi - \upsilon_i) - \sigma'_{\upsilon_i x}\Sigma^{*-1}_{k,2}x^k_d].$$

Given that in the log-normal setup $\omega_{\upsilon_i|x}$ is constant and $\varsigma_{\upsilon_i|x}$ is linear in the log of x_i, then $E(V_i|X_1, \ldots, X_N)$ is monotonically increasing in X_i.

For a proof of the existence of an increasing-strategy Bayesian-Nash equilibrium see Milgrom and Weber (1982) theorem 10 and Hong and Shum (2003, pp. 352). Specifically, they show that if all bidders $j \neq i$ follow their equilibrium strategies $\beta_j^k(\cdot)$, bidder i's best response is to play $\beta_i^k(\cdot)$ because this guarantees that bidder i will win the auction if and only if his expected net payoff is positive conditional on winning.

1.7.2 *Mean, variances and covariances of values and signals*

Noting that $v_i = E(v_i) = E(a_i + v) = m + \bar{a}_i$, then $v = (v_1, \ldots, v_N)' = (m + \bar{a}_1, \ldots, m + \bar{a}_N)'$. Similarly, $E(x_i) = E(v_i + s_i\xi_i) = E(v_i) = m + \bar{a}_i$, so $\Psi = E(x) = (m + \bar{a}_1, \ldots, m + \bar{a}_N)'$. Also,

$$Var(v_i) = Var(a_i + v) = Var(a_i) + Var(v) + 2Cov(a_i, v) = r_0^2 + t_i^2$$

and

$$Cov(v_i, v_j) = E(v_i v_j) - E(v_i)E(v_j) = E(v^2) - [E(v)]^2 = Var(v) = r_0^2$$

for all $i, j \in N$ and $i \neq j$, so

$$\sigma_v^2 = \begin{pmatrix} r_0^2 + t_1^2 & \cdots & r_0^2 \\ \vdots & \ddots & \vdots \\ r_0^2 & \cdots & r_0^2 + t_N^2 \end{pmatrix}.$$

In addition,

$$Cov(v_i, x_i) = E[v_i(v_i + s_i\xi_i)] - E(v_i)E(x_i) = E(v_i^2) - [E(v_i)]^2 = Var(v_i)$$

and

$$Cov(v_i, x_j) = E[v_i(v_j + s_j\xi_j)] - E(v_i)E(x_j) = E(v_i v_j) - E(v_i)E(v_j) = Cov(v_i, v_j)$$

for all $i, j \in N$ and $i \neq j$. As a consequence,

$$\sigma_{v_i\Psi} = \begin{pmatrix} r_0^2 + t_1^2 & \cdots & r_0^2 \\ \vdots & \ddots & \vdots \\ r_0^2 & \cdots & r_0^2 + t_N^2 \end{pmatrix}.$$

Finally, since

$$Var(x_i) = Var(v_i + s_i\xi_i) = Var(v_i) + s_i^2 Var(\xi_i) + 2s_i Cov(v_i, \xi_i) = r_0^2 + t_i^2 + s_i^2$$

and

$$Cov(x_i, x_j) = E(x_i x_j) - E(x_i)E(x_j) = E(v_i v_j) - E(v_i)E(v_j) = Cov(v_i, v_j)$$

for all $i, j \in N$ and $i \neq j$, we have that

$$\Sigma^* = \begin{pmatrix} r_0^2 + t_1^2 + s_1^2 & \cdots & r_0^2 \\ \vdots & \ddots & \vdots \\ r_0^2 & \cdots & r_0^2 + t_N^2 + s_N^2 \end{pmatrix}.$$

1.7.3 *Difference between loss aversion and expected utility*

When $\lambda_i = 1$, one gets the standard risk neutral case (see Hong and Shum (2003)), so

$$\exp[\frac{1}{2}Var(v_i^k|\Upsilon_i^k)]\exp[E(v_i^k|\Upsilon_i^k)] = \exp[\ln(\beta_i^k(\cdot)]$$

or equivalently

$$E(v_i^k|\Upsilon_i^k) = \ln[\beta_i^k(\cdot)] - \frac{1}{2}Var(v_i^k|\Upsilon_i^k),$$

with $x_r^k = (\sigma'_{\upsilon_i x}\Sigma_{k,1}^{*-1})^{-1}[E(v_i^k|\Upsilon_i^k) + (\sigma'_{\upsilon_i x}\Sigma^{*-1}\Psi - \upsilon_i) - \sigma'_{\upsilon_i x}\Sigma_{k,2}^{*-1}x_d^k]$.

Hence, the difference between loss aversion (LA) and expected utility (EU) in this ascending model is simply:

$$E^{LA}(v_i^k|\Upsilon_i^k) - E^{EU}(v_i^k|\Upsilon_i^k) = \eta_i.$$

Moreover, since $x_r^k = (\sigma'_{\upsilon_i x}\Sigma_{k,1}^{*-1})^{-1}[E(v_i^k|\Upsilon_i^k) + (\sigma'_{\upsilon_i x}\Sigma^{*-1}\Psi - \upsilon_i) - \sigma'_{\upsilon_i x}\Sigma_{k,2}^{*-1}x_d^k]$, then at round 0,

$$x_r^{0,PR} - x_r^{0,EU} = (\sigma'_{\upsilon_i x}\Sigma_{k,1}^{*-1})^{-1}\eta_i,$$

while in subsequent rounds,

$$x_r^{k,PR} - x_r^{k,EU} = \left(\sigma'_{\upsilon_i x}\Sigma_{k,1}^{*-1}\right)^{-1}(\eta_i - \sigma'_{\upsilon_i x}\Sigma_{k,2}^{*-1}x_d^{k,PR} + \sigma'_{\upsilon_i x}\Sigma_{k,2}^{*-1}x_d^{k,EU}).$$

1.7.4 *Special cases*

There are three important special cases of the model under prospect theory proposed in section 1.3: the independent private value model, the pure common value model and Wilson's (1998) model. In addition, any heterogeneous model may simplify to a fully homogeneous one.

Independent private value model

In this model there is no correlation in the valuations, so that $V_i = A_i$, which effectively requires that $V = 1$ ($m = r_0^2 = 0$), implying that the only information bidders care about is their own valuation. The log-bidding functions for each bidder in the initial round only depend on their own private signal. As the auction progresses, the bidding functions do not change in subsequent rounds. Therefore, Vickrey (1961) revenue equivalence theorem result applies. Specifically, Vickrey (1961) showed that in a sealed-bid second price independent private value auction, the bidders' optimal strategies are to truthfully bid their valuations. As the valuation of each bidder is independent of the others, observing someone else's valuation has no impact on the valuation of anyone else, making the bids independent of the number of bidders participating in the auction. Therefore, as all bidders will drop out when the price reaches their privately known values, the outcome is Pareto optimal because the bidder with the highest value will win the item.

Pure common value model

In this special case, $V_i = V$, which requires that $A_i = 1$ for all i ($\bar{a}_i = t_i^2 = 0$). As in the independent private value model, the log-bidding functions for each bidder in the initial round only depend on their own private signal. However, as the auction progresses, bidders lower their bids. Intuitively, any information possessed by the bidders who drop out significantly influences the beliefs of the other bidders because the value of the object is the same across bidders.

Wilson auction model

Wilson (1998) allowed each bidder to observe two signals: his private component A_i as well as his noisy estimate of the common component E_i. In contrast, the model in section 1.3 only allows bidder i to obsere the composite signal $X_i = A_i \times E_i$. Nevertheless, given that he made a diffuse prior assumption on the distribution of the common value component, in practice, one can achieve the same with $r_0^2 = \infty$. As usual, the log-bidding functions for each bidder at the initial round only depend on their own private signals, but as the auction progresses, the effects of the private signal on the log-bidding functions of the remaining bidders is greater than in both the pure common value and independent private value models. The reason is that the diffuse nature of the prior makes bidders pay more attention to all the signals they observe.

Fully homogeneous model

In the fully homogeneous model, all the parameters are common, implying that they do not depend on the bidders' characteristics, i.e. $\theta \equiv (\bar{a}, m, t^2, r_0^2, s^2)$. In this case, the bidder with the lowest signal will be the first one to drop out in round 0 because all bidders are homogeneous. Given that the winner is the bidder with the highest bid, in a homogeneous set up this corresponds to the one with the highest signal.

1.7.5 *Winner's curse*

In first and second price sealed bid common value auctions, there may exist winner's curse, i.e. overpaying due to incomplete information. Suppose there are 3 bidders and the item for auction has an actual value of \$5. Assume bidder A bids \$2, bidder B bids \$6 and bidder C \$9. Even though bidder C won the auction, he ended up overpaying by \$4. If bidders take this problem into consideration, they should shade their bids, leading the average bid to decrease with the number of bidders, as in Athey and Haile (2002) and Bajari and Hortaçsu (2003). However, ascending auctions have the unique feature of "information transparency", so bidders can make inferences about the private information possessed by the bidders who have dropped out. As shown in Milgrom and Weber (1982), this feature reduces the effects of the winner's curse, allowing bidders to bid more aggressively than in a sealed-bid auctions.

To shed some light on the existence of winner's curse in *Storage Wars*, the following table reports the average profit for the regular bidders:

Winners	Estimate	Std. Error
Barry	920.78	2407.01
Darrell	9473.71	2676.39
Dave	3949.31	2779.37
Jarrod	1721.38	2481.08
Anonymous	915.91	5984.59

Notes: *Indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

As in Hong and Shum (2002) and Bajari and Hortaçsu (2003), I find that bidders took into account the possibility of overpaying and shaded their bids in order to avoid winner's curse. These results suggest that professional bidder behavior is based on previous experiences, which is in line with the experimental evidence in Kagel and Levin (1986).

1.7.6 *Calculating the likelihood of baseline model*

Continuous component

Using the log-bidding function in Section 1.3.2, the bid functions of bidders dropping out in round k will be given by (1.7). Let

$$F = \begin{pmatrix} \frac{C_N^0}{A_N^0} - \eta_N & \cdots & \frac{C_2^{N-2}}{A_2^{N-2}} - \eta_2 \end{pmatrix}$$

be an $(N-1) \times 1$ vector,

$$\mathcal{G}_i = \begin{pmatrix} \underbrace{0, \ldots, 0}_{N-i-2} & 1/\mathcal{A}_{N-i}^i & \mathcal{D}_{N-i}^i/\mathcal{A}_{N-i}^i \end{pmatrix}$$

a $1 \times (N-1)$ vector and

$$\mathcal{G} = \begin{pmatrix} \mathcal{G}_0' & \cdots & \mathcal{G}_{N-2}' \end{pmatrix}$$

an $(N-1) \times (N-1)$ matrix. Thus, the vector of dropout bids can be written as

$$\mathcal{P} = \mathcal{G}\,(x_2, \ldots, x_N)' + \mathcal{F}. \tag{1.10}$$

Let $\psi_2(\theta)$ be the $N-1$ subvector of Ψ and $\Sigma_2^*(\theta)$ the $(N-1) \times (N-1)$ submatrix of Σ^* corresponding to bidders $2, \ldots, N$. Then, equation (1.10) implies that the mean and variance of the vector of dropout bids will be

$$\left.\begin{aligned} \mu_p(\theta) &= \mathcal{F}(\theta) + \mathcal{G}(\theta)\psi_2(\theta) \\ \Sigma_p(\theta) &= \mathcal{G}(\theta)\Sigma_2^*(\theta)\mathcal{G}(\theta)' \end{aligned}\right\}.$$

Therefore, the continuos part of the $(N-1)$-variate normal log-likelihood function for a given auction is

$$\log f(\mathcal{P};\theta) = -\frac{1}{2}(N-1)\log(2\pi) - \frac{1}{2}\log(|\Sigma_p(\theta)|) - \frac{1}{2}\left\{[\mathcal{P} - \mu_p(\theta)]'\Sigma_p(\theta)^{-1}[\mathcal{P} - \mu_p(\theta)]\right\}.$$

Characterization of $T_2(\theta)$ and its probability

In an ascending auction, one does not observe the winner's dropout bid, only the price at which the second highest bidder stops. As a result, the signal of the winning bidder is constrained to a region $\mathcal{T}_2(x_2, \ldots, x_N;\theta) \subset R^1$. Hong and Shum (2003) show that the set $\mathcal{T}_2[\mathcal{G}^{-1}(\mathcal{P} - \mathcal{F});\theta]$ consist of the following conditions:

$$\{x_1 : b_1^l(x_1; x_d^l, \theta) \geq p_l, \text{ for all } l = 0, \ldots, N-2\}.$$

This implies that for any dropout order, the winning bidder will never regret having remained active in all prior rounds. However, given the ascending nature of the auction, the only binding constraint will be

$$b_1^{N-2}(x_1; x_d^{N-2}, \theta) \geq p_{N-2}. \tag{1.11}$$

Unfortunately, there is a mistake in the expression for the probability of $\mathcal{T}_2$ after the formula (24) that Hong and Shum (2003) provide. Specifically, they seem to have used unconditional moments when they should have used conditional ones instead because $\Pr\{\mathcal{T}_2[\mathcal{G}^{-1}(\mathcal{P} - \mathcal{F}); \theta]\}$ denotes the probability that $x_1 \in \mathcal{T}_2(\theta)$ conditional on $\mathcal{P}$.

To illustrate the calculation, consider an auction with $N = 3$ bidders. Without loss of generality, suppose bidder 3 had the lowest bid in round 0, so at round 1 only bidders 1 and 2 remain active. Then,

$$b_1^1(x_1; x_3, \theta) \geq p_1,$$

which can then be simplified to

$$x_1 \geq \mathcal{A}_1^1 p_1 - \mathcal{C}_1^1 - \mathcal{D}_1^1 x_3 + \mathcal{A}_1^1 \eta_1.$$

Therefore,

$$\Pr\{\mathcal{T}_2[\mathcal{G}^{-1}(\mathcal{P} - \mathcal{F})|\theta]\} = \Pr\left[\frac{x_1 - E(x_1|x_2, x_3)}{\sqrt{Var(x_1|x_2, x_3)}} \geq \frac{\mathcal{A}_1^1 p_1 - \mathcal{C}_1^1 - \mathcal{D}_1^1 x_3 + \mathcal{A}_1^1 \eta_1 - E(x_1|x_2, x_3)}{\sqrt{Var(x_1|x_2, x_3)}}\right]$$

or equivalently

$$\Pr\{\mathcal{T}_2[\mathcal{G}^{-1}(\mathcal{P} - \mathcal{F})|\theta]\} = \Phi\left[\frac{E(x_1|x_2, x_3) + \mathcal{C}_1^1 + \mathcal{D}_1^1 x_3 - \mathcal{A}_1^1 p_1 - \mathcal{A}_1^1 \eta_1}{\sqrt{Var(x_1|x_2, x_3)}}\right].$$

To obtain $E(x_1|x_d) = \bar{\Psi}$ and $Var(x_1|x_d) = \bar{\Sigma}^*$, first partition the vector x as

$$x = [x_1 \ \underbrace{(x_2, x_3)}_{N-1}]',$$

and then partition Ψ and Σ^* accordingly:

$$\Psi = [\Psi_1 \ \underbrace{\Psi_d}_{N-1}]' \ \text{ and } \ \Sigma^* = \begin{pmatrix} \Sigma_{11}^* & \overbrace{\Sigma_{1d}^*}^{1\times(N-1)} \\ \underbrace{\Sigma_{d1}^*}_{(N-1)\times 1} & \underbrace{\Sigma_{dd}^*}_{(N-1)\times(N-1)} \end{pmatrix}.$$

Then, the distribution of x_1 conditional on (x_2, x_3) is multivariate normal $x_1|x_2, x_3 \sim N(\bar{\Psi}, \bar{\Sigma}^*)$, where $\bar{\Psi} = \Psi_1 + \Sigma_{1d}^* \left(\Sigma_{dd}^*\right)^{-1} [(x_2, \ldots, x_{N+q}) - \Psi_d]$ and $\bar{\Sigma}^* = \Sigma_{11}^* - \Sigma_{1d}^* \left(\Sigma_{dd}^*\right)^{-1} \Sigma_{d1}^*$.

Characterization of $\Pr[T_1(\theta);\theta]$

For the dropout to occur in the correct order (CO), it must be the case that

$$b_i^k(x_i; x_d^k, \theta) \geq b_{N-k}^k(x_{N-k}; x_d^k, \theta) = p_k, \text{ for all } k \text{ and } i = 0, \ldots, N-k-1.$$

The truncation region $\mathcal{T}_1(\theta)$ for a given value of θ is defined as the values of the log-signals such that CO is satisfied. More formally,

$$\mathcal{T}_1(\theta) = \{x_1, \ldots, x_N : CO \text{ is satisfied}|\theta\}.$$

Given the ascending nature of the auction and that the log-bidding functions for rounds k and $k-1$ intersect when they are equal, Hong and Shum (2003) show that the CO condition can be simplified to the following $N-1$ inequalities

$$b_{N-k-1}^k(x_{N-k-1}; x_d^k, \theta) \geq b_{N-k}^k(x_{N-k}; x_d^k, \theta), \text{ for all } k = 0, \ldots, N-2,$$

which implies that the log-bidding functions of the bidders remaining in round k have to be greater than the log-bidding functions of all the ones who have dropped out.

To illustrate the calculations for $\Pr[T_1(\theta);\theta]$, suppose that, for example, $N = 3$. The only binding constraints are:

$$\left.\begin{array}{l} b_2^0(x_2;\theta) \geq b_3^0(x_3;\theta) \\ b_1^1(x_1, x_3;\theta) \geq b_2^1(x_2, x_3;\theta) \end{array}\right\}$$

which can be written in matrix form as

$$\underbrace{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}_{Z} \geq \underbrace{\begin{bmatrix} \frac{C_3^0}{A_3^0} - \frac{C_2^0}{A_2^0} + (\eta_2 - \eta_3) \\ \frac{C_2^1}{A_2^1} - \frac{C_1^1}{A_1^1} + (\eta_1 - \eta_2) \end{bmatrix}}_{h} + \underbrace{\begin{bmatrix} 0 & -\frac{1}{A_2^0} & \frac{1}{A_3^0} \\ -\frac{1}{A_1^1} & \frac{1}{A_2^1} & \left(\frac{D_2^1}{A_2^1} - \frac{D_1^1}{A_1^1}\right) \end{bmatrix}}_{H} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}}_{x},$$

The probability that $Z \leq 0$ is simply a multivariate normal cdf with $E(Z) = h + H\Psi$ and $V(Z) = H\Sigma^* H'$ because $x \sim N(\Psi, \Sigma^*)$. To calculate this multivariate normal cdf, I use a numerical quadrature procedure for bivariate and trivariate distributions, and a quasi-Monte Carlo integration algorithm for four or more dimensions (see Matlab (2019) *mvncdf* entry for more details).

As an aside, it is worth mentioning that the CO condition in $\Pr[\mathcal{T}_1(\theta)|\theta]$ in the fully homogeneous case (see section 1.7.4) implies that the log-signal of the winner has to be greater than the log-signal of the second highest bidder, and similarly the log-singal of the third highest bidder, etc. For example, when there are only three bidders,

$$\Pr[\mathcal{T}_1(\theta)|\theta] = \Pr(x_1 \geq x_2; x_2 \geq x_3|\theta) = \Pr(z_1 \geq 0; z_2 \geq 0|\theta),$$

where $z_1 = x_1 - x_2$ and $z_2 = x_2 - x_3$. But notice that this is the probability that a bivariate normal with zero means, unit variances and some correlation coefficient between z_1 and z_2 ($\rho_{z_1 z_2}$) lies in the first quadrant. In this case, it is easy to prove that

$$\Pr[\mathcal{T}_1(\theta)|\theta] = 1/N!$$

because there are $N!$ possible orderings, which are all equally likely in the fully homogeneous case. Consequently, $\Pr[\mathcal{T}_1(\theta)|\theta]$ does not depend on the model parameters.

1.7.7 *Calculating the likelihood with active non-bidding participants*

Continuous component

Define

$$\tilde{\mathcal{F}} = \begin{pmatrix} \frac{\mathcal{C}_{N+q}^{-1}}{\mathcal{A}_{N+q}^{-1}} - \eta_{N+q} & \cdots & \frac{\mathcal{C}_{N+1}^{-1}}{\mathcal{A}_{N+1}^{-1}} - \eta_{N+1} & \frac{\mathcal{C}_{N}^{0}}{\mathcal{A}_{N}^{0}} - \eta_{N} & \cdots & \frac{\mathcal{C}_{2}^{N-2}}{\mathcal{A}_{2}^{N-2}} - \eta_{2} \end{pmatrix}$$

as an $(N+q-1) \times 1$ vector, with q being the number of non-bidding participants and N the number of active bidders. Similarly, let

$$\begin{aligned} \tilde{\mathcal{G}}_j &= \begin{pmatrix} \underbrace{0,\ldots,0}_{N-j-2} & 1/\mathcal{A}_{N-j}^{-1} & \underbrace{0,\ldots,0}_{q+j} \end{pmatrix}, \text{ for } j = -q,\ldots,-1 \\ \tilde{\mathcal{G}}_i &= \begin{pmatrix} \underbrace{0,\ldots,0}_{N-i-2} & 1/\mathcal{A}_{N-i}^{i} & \underbrace{\mathcal{D}_{N-i}^{i}/\mathcal{A}_{N-i}^{i}}_{q+i} \end{pmatrix}, \text{ for } i = 0,\ldots,N-2. \end{aligned}$$

denote two $1 \times (N+q-1)$ vectors and

$$\tilde{\mathcal{G}} = \begin{pmatrix} \tilde{\mathcal{G}}'_{-q} & \cdots & \tilde{\mathcal{G}}'_{-1} & \tilde{\mathcal{G}}'_{0} & \cdots & \tilde{\mathcal{G}}'_{N-2} \end{pmatrix}$$

an $(N+q-1) \times (N+q-1)$ matrix. As before, the vector of dropout bids can be written as

$$\tilde{\mathcal{P}} = \tilde{\mathcal{G}}\,(x_2,\ldots,x_{N+q})' + \tilde{\mathcal{F}}. \tag{1.12}$$

This equation describes the mapping from the unobserved log-signals

$$x_{dr} \equiv (x_2,\ldots,x_N, x_{N+1},\ldots,x_{N+q})'$$

to the observed log-bids $\tilde{\mathcal{P}} = (\underbrace{p_{-1}, \ldots, p_{-1}}_{q}, p_0, \ldots, p_{N-2})'$.

Let $\tilde{\psi}_2(\theta)$ be the $N+q-1$ subvector of Ψ and $\tilde{\Sigma}_2^*(\theta)$ the $(N+q-1)\times(N+q-1)$ submatrix of Σ^* corresponding to the signals of bidders $2, \ldots, N+q$. Then, equation (1.12) implies that the mean and variance of the vector of dropout bids will be

$$\left.\begin{array}{l} \tilde{\mu}_p(\theta) = \tilde{\mathcal{F}}(\theta) + \tilde{\mathcal{G}}(\theta)\tilde{\psi}_2(\theta) \\ \tilde{\Sigma}_p(\theta) = \tilde{\mathcal{G}}(\theta)\tilde{\Sigma}_2^*(\theta)\tilde{\mathcal{G}}(\theta)' \end{array}\right\}.$$

Similarly, partition the price vector $\tilde{\mathcal{P}}$ as:

$$\tilde{\mathcal{P}} = (\underbrace{(p_{-1}, \ldots, p_{-1})}_{q} \;\; \underbrace{(p_0, \ldots, p_{N-2})}_{N-1})'$$

and then partition $\tilde{\mu}_p(\theta)$ and $\tilde{\Sigma}_p(\theta)$ accordingly:

$$\tilde{\mu}_p(\theta) = (\underbrace{\tilde{\mu}_{p,1}}_{q} \;\; \underbrace{\tilde{\mu}_{p,2}}_{N-1})' \;\text{ and }\; \tilde{\Sigma}_p(\theta) = \begin{pmatrix} \overbrace{\tilde{\Sigma}_{p,11}}^{q\times q} & \overbrace{\tilde{\Sigma}_{p,12}}^{q\times(N-1)} \\ \underbrace{\tilde{\Sigma}_{p,21}}_{(N-1)\times q} & \underbrace{\tilde{\Sigma}_{p,22}}_{(N-1)\times(N-1)} \end{pmatrix}.$$

Then, the distribution of $(p_0, \ldots, p_{N-2})$ conditional on $(p_{-1}, \ldots, p_{-1})$ is multivariate normal $[(p_0, \ldots, p_{N-2}) \,|\, (p_{-1}, \ldots, p_{-1})] \sim N\left(\bar{\mu}_p, \bar{\Sigma}_p\right)$, where $\bar{\mu}_p = \tilde{\mu}_{p,2} + \tilde{\Sigma}_{p,21}\tilde{\Sigma}_{p,11}^{-1}\left[(p_{-1}, \ldots, p_{-1}) - \tilde{\mu}_{p,1}\right]$ and $\bar{\Sigma}_p = \tilde{\Sigma}_{p,22} - \tilde{\Sigma}_{p,21}\tilde{\Sigma}_{p,11}^{-1}\tilde{\Sigma}_{p,12}$.

Therefore, the continuos part of the $(N-1-q)$-variate normal log-likelihood function for a given auction conditional on the initial dropout bidders is

$$\log f(\mathcal{P}_a|P_{-1}, \theta) = -\frac{1}{2}(N-1-q)\log(2\pi) - \frac{1}{2}\log(|\bar{\Sigma}_p(\theta)|) - \frac{1}{2}\left\{[\mathcal{P}_a - \bar{\mu}_p(\theta)]'\bar{\Sigma}_p(\theta)^{-1}[\mathcal{P}_a - \bar{\mu}_p(\theta)]\right\},$$

where $\mathcal{P}_a = (p_0, \ldots, p_{N-2})'$.

Characterization of $T_2(\theta)$ **and** $\Pr[T_1(\theta); \theta|P_{-1}]$

In this case, the probability of $\mathcal{T}_2$ will be the same as (1.11). To illustrate how the $\Pr[T_1(\theta); \theta]$ looks like in this context suppose that, for example, $N = 3$ and $q = 2$.

At round -1, bidders 5 and 4 drop out simultaneously at price p_{-1}. Therefore, the only binding constraints will be:

$$\left.\begin{array}{l} b_3^{-1}\left(x_3|x_4,x_5\right) \geq b_4^{-1}\left(x_4|x_4,x_5\right) \\ b_2^{0}\left(x_2|x_4,x_5\right) \geq b_3^{0}\left(x_3|x_4,x_5\right) \\ b_1^{1}\left(x_1;x_d|x_4,x_5\right) \geq b_2^{1}\left(x_2;x_d|x_4,x_5\right) \end{array}\right\}$$

which can be written in matrix form as

$$\underbrace{\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}}_{Z} \geq \underbrace{\begin{bmatrix} \frac{\mathcal{C}_4^{-1}}{\mathcal{A}_4^{-1}} - \frac{\mathcal{C}_3^{-1}}{\mathcal{A}_3^{-1}} + (\eta_3 - \eta_4) \\ \frac{\mathcal{C}_3^{0}}{\mathcal{A}_3^{0}} - \frac{\mathcal{C}_2^{0}}{\mathcal{A}_2^{0}} + (\eta_2 - \eta_3) \\ \frac{\mathcal{C}_2^{1}}{\mathcal{A}_2^{1}} - \frac{\mathcal{C}_1^{1}}{\mathcal{A}_1^{1}} + (\eta_1 - \eta_2) \end{bmatrix} + \begin{bmatrix} \frac{1}{\mathcal{A}_4^{-1}} & 0 \\ \left(\frac{\mathcal{D}_{3,1}^{0}}{\mathcal{A}_3^{0}} - \frac{\mathcal{D}_{2,1}^{0}}{\mathcal{A}_2^{0}}\right) & \left(\frac{\mathcal{D}_{3,2}^{0}}{\mathcal{A}_3^{0}} - \frac{\mathcal{D}_{2,2}^{0}}{\mathcal{A}_2^{0}}\right) \\ \left(\frac{\mathcal{D}_{2,2}^{1}}{\mathcal{A}_2^{1}} - \frac{\mathcal{D}_{1,2}^{1}}{\mathcal{A}_1^{1}}\right) & \left(\frac{\mathcal{D}_{2,3}^{1}}{\mathcal{A}_2^{1}} - \frac{\mathcal{D}_{1,3}^{1}}{\mathcal{A}_1^{1}}\right) \end{bmatrix} \underbrace{\begin{pmatrix} x_4 \\ x_5 \end{pmatrix}}_{x_b}}_{h}$$

$$+ \underbrace{\begin{bmatrix} 0 & 0 & -\frac{1}{\mathcal{A}_3^{-1}} \\ 0 & -\frac{1}{\mathcal{A}_2^{0}} & \frac{1}{\mathcal{A}_3^{0}} \\ -\frac{1}{\mathcal{A}_1^{1}} & \frac{1}{\mathcal{A}_2^{1}} & \left(\frac{\mathcal{D}_{2,1}^{1}}{\mathcal{A}_2^{1}} - \frac{\mathcal{D}_{1,1}^{1}}{\mathcal{A}_1^{1}}\right) \end{bmatrix}}_{H} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}}_{x_a},$$

Note that this characterization is equivalent to $b_3^{-1}\left(x_3|x_4,x_5\right) \geq b_5^{-1}\left(x_5|x_4,x_5\right)$.

Then, partition the vector x as

$$x = \left(x_a', x_b'\right)' = [\underbrace{(x_1,\ldots,x_N)}_{N} \; \underbrace{(x_{N+1},\ldots,x_{N+q})}_{q}]',$$

and then partition Ψ and Σ^* accordingly:

$$\Psi = (\underbrace{\Psi_a}_{N} \;\; \underbrace{\Psi_b}_{q})' \quad \text{and} \quad \Sigma^* = \begin{pmatrix} \overbrace{\Sigma_{aa}^*}^{N\times N} & \overbrace{\Sigma_{ab}^*}^{N\times q} \\ \underbrace{\Sigma_{ba}^*}_{q\times N} & \underbrace{\Sigma_{bb}^*}_{q\times q} \end{pmatrix}.$$

The distribution of x_a conditional on x_b is multivariate normal $x_a|x_b \sim N(\hat{\Psi},\hat{\Sigma}^*)$, where $\hat{\Psi} = \Psi_a + \Sigma_{ab}^*\left(\Sigma_{bb}^*\right)^{-1}[(x_{N+1},\ldots,x_{N+q}) - \Psi_b]$ and $\hat{\Sigma}^* = \Sigma_{aa}^* - \Sigma_{ab}^*\left(\Sigma_{bb}^*\right)^{-1}\Sigma_{ba}^*$.

The probability that $Z \leq 0$ conditional on x_b is simply a multivariate normal cdf with $E[Z|\left(x_{N+1},\ldots,x_{N+q}\right)] = h + H\tilde{\Psi}$ and $V[Z|\left(x_{N+1},\ldots,x_{N+q}\right)] = H\tilde{\Sigma}^*H'$.

1.8 Tables Chapter 1

Table 1.1: Auctioneer Behavior

	Estimate	Std. Error
HHI^{**}	0.012	0.005
$SIZE^{***}$	0.472	0.103
Constant	1.829	0.395

Notes: Multiple regression of (log) opening bid. HHI captures the median household income of the municipality where the locker is located in the State of California, $SIZE$ is a variable that measures the size of the locker (small (1), medium (2) or large (3)). Additionally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

Table 1.2: Summary Statistics

Variable	#Obs.	Season 1	Season 2	Season 3
Auction characteristics				
Small locker	100	23	37	40
Medium locker	117	27	54	36
Large locker	37	9	12	16
Average HHI	78	57641	61473	58750
Average *Ex-post*	250	4797	3954	6319
Average Profit	250	3949	2282	4821
Number of auctions	254	59	103	92
Number of bidders per auction				
$N = 2$	14	3	6	5
$N = 3$	50	9	13	28
$N = 4$	72	20	29	23
$N = 5$	63	13	34	16
$N = 6$	36	8	16	12
$N = 7$	19	6	5	8

Notes: HHI denotes the median household income of the municipality where the locker is located in the State of California, *Ex-post* denotes the ex-post value of the locker and profit denotes the difference between the ex-post value of the locker and the winner's bid.

Table 1.3: Bidder's Frequency Participation

	# Obs.	First Bidder	J	Dr	Dv	J-Dr	J-Dv	Dr-Dv	J-Dr-Dv
Barry	139	66	82	86	76	52	44	50	29

	# Obs.	First Bidder	J	B	Dv	J-B	J-Dv	B-Dv	J-B-Dv
Darrell	161	25	96	86	98	52	57	50	29

	# Obs.	First Bidder	J	Dr	B	J-Dr	J-B	Dr-B	J-Dr-B
Dave	151	10	99	98	76	57	44	50	29

	# Obs.	First Bidder	B	Dr	Dv	B-Dr	B-Dv	Dr-Dv	B-Dr-Dv
Jarrod	165	31	82	96	99	52	44	57	29

Notes: The four main bidders are Barry "B", Darrell "Dr", Dave "Dv" and Jarrod "Jr". Additionally, "J-Dv" means that Jarrod and Dave were the only two main bidders out of the four who were active bidding participants, i.e. they participated in the auction.

Table 1.4: Mean Common Value

	Estimate	Std. Error
HHI^{**}	0.012	0.005
$SIZE^{***}$	0.368	0.123
Constant	6.242	0.408

Notes: Multiple regression of (log) *Ex-post* value. HHI captures the median household income of the municipality where the locker is located in the State of California and $SIZE$ is a variable that measures the size of the locker (small (1), medium (2) or large (3)). Additionally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

Table 1.5: Maximum Likelihood Estimates Baseline Model

	Estimate	p-value
β_1^{***}	0.381	0
β_2^{**}	0.009	0.01
δ_0	0.014	-
α_0	5.534	-
α_1^{***}	-0.185	0
α_2^{***}	0.198	0
α_3^{*}	0.101	0.06
α_4	0.039	0.51
τ_0	-1.081	-
γ_0	0.371	-
γ_1	-0.059	0.85
γ_2	0.053	0.77
γ_3^{***}	0.923	0
γ_4^{*}	-0.717	0.09

Notes: The mean of the common value component for a given auction is $m = \beta_0 + \beta_1 SIZE + \beta_2 HHI$, where $SIZE$ is a variable that measures the size of the locker (small (1), medium (2) or large (3)) and HHI captures the median household income of the municipality where the locker is located in the State of California. Additionally, the mean of the private value component of the four main bidders (Barry "Ba", Darrell "Dr", Dave "Dv" and Jarrod "Jr"), as well as of the other active bidders whose identity is not shown publicly, is $\bar{a} = (\alpha_0 + \alpha_1 Ba, \alpha_0 + \alpha_2 Dr, \alpha_0 + \alpha_3 Dv, \alpha_0 + \alpha_4 Jr, \alpha_0, \ldots, \alpha_0)$, where Ba, Dr, Dv and Jr are mutually exclusive dummy variables. Furthermore, the variance of the common and private value component is modelled as $r_0^2 = \exp(\delta_0)$ and $t^2 = \exp(\tau_0)$, respectively, while the variance of the noise for each of the bidder's signals is $s^2 = \exp(\gamma_0 + \gamma_1 Ba, \gamma_0 + \gamma_2 Dr, \gamma_0 + \gamma_3 Dv, \gamma_0 + \gamma_4 Jr, \gamma_0, \ldots, \gamma_0)$. p-values correspond to the likelihood ratio. Finally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

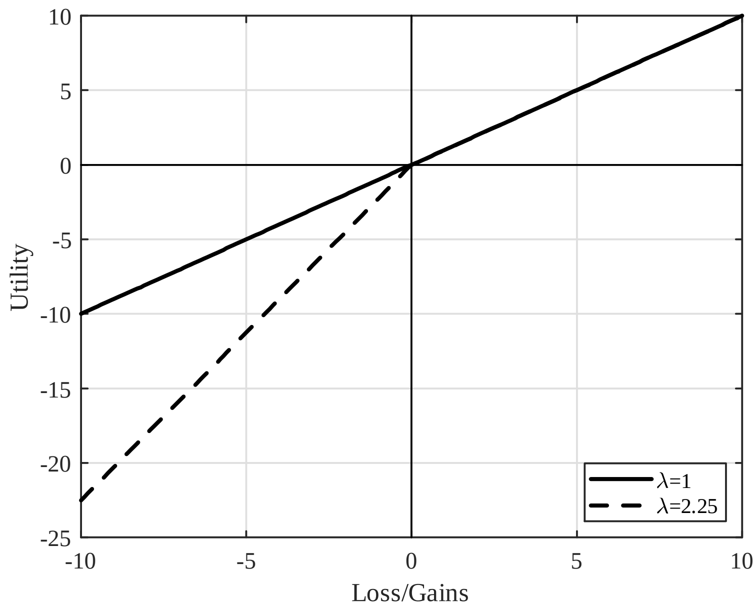
Table 1.6: Maximum Likelihood Estimates With Active Non-Bidding Participants

	Estimate	p-value
β_1^{***}	0.291	0
β_2^{***}	0.009	0
δ_0	-0.002	-
α_0	5.926	-
α_1^{***}	-0.439	0
α_2	-0.007	0.55
α_3^{***}	-0.126	0
α_4	0.001	0.59
τ_0	0.001	-
γ_0	1.628	-
γ_1^{*}	-0.427	0.08
γ_2	-0.022	0.79
γ_3^{***}	1.024	0
γ_4	0.058	0.76

Notes: The mean of the common value component for a given auction is $m = \beta_0 + \beta_1 SIZE + \beta_2 HHI$, where $SIZE$ is a variable that measures the size of the locker (small (1), medium (2) or large (3)) and HHI captures the median household income of the municipality where the locker is located in the State of California. Additionally, the mean of the private value component of the four main bidders (Barry "Ba", Darrell "Dr", Dave "Dv" and Jarrod "Jr"), as well as of the other active bidders whose identity is not shown publicly, is $\bar{a} = (\alpha_0 + \alpha_1 Ba, \alpha_0 + \alpha_2 Dr, \alpha_0 + \alpha_3 Dv, \alpha_0 + \alpha_4 Jr, \alpha_0, \ldots, \alpha_0)$, where Ba, Dr, Dv and Jr are mutually exclusive dummy variables. Furthermore, the variance of the common and private value component is modelled as $r_0^2 = \exp(\delta_0)$ and $t^2 = \exp(\tau_0)$, respectively, while the variance of the noise for each of the bidder's signals is $s^2 = \exp(\gamma_0 + \gamma_1 Ba, \gamma_0 + \gamma_2 Dr, \gamma_0 + \gamma_3 Dv, \gamma_0 + \gamma_4 Jr, \gamma_0, \ldots, \gamma_0)$. p-values correspond to the likelihood ratio. Finally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

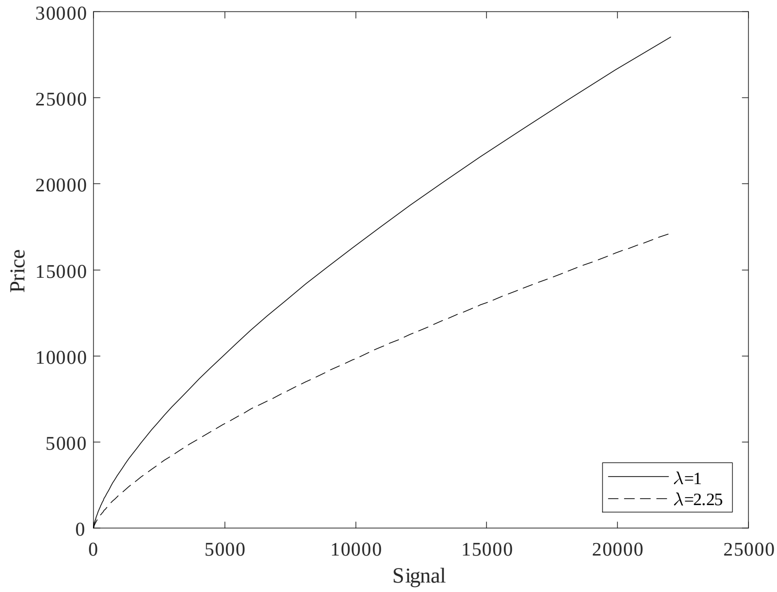
1.9 Graphs Chapter 1

Figure 1.1: Loss Aversion Utility Function



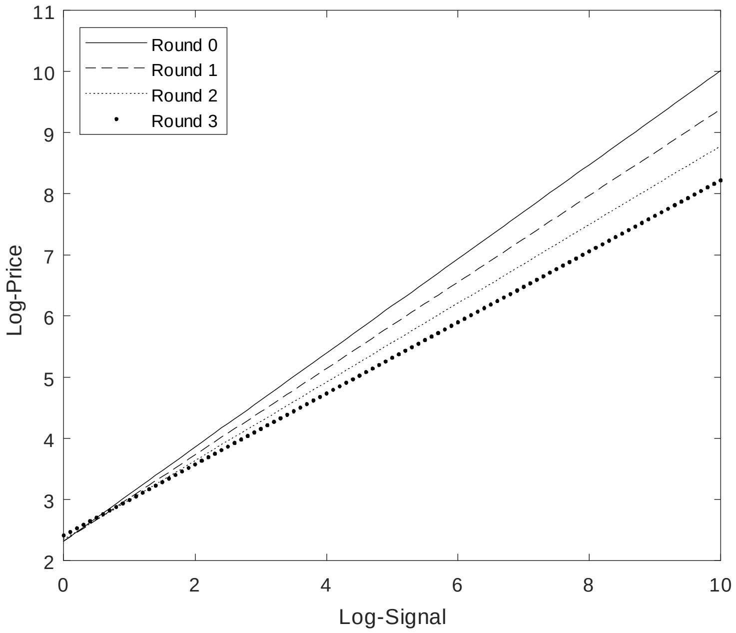
Notes: This graph displays the shape of the utility function (1.1) plotted against gains and losses for $\lambda = 1$ (risk neutrality) and $\lambda = 2.25$ (loss aversion), with the marginal utility of losses being λ times the marginal utility of gains.

Figure 1.2: Bidding Functions



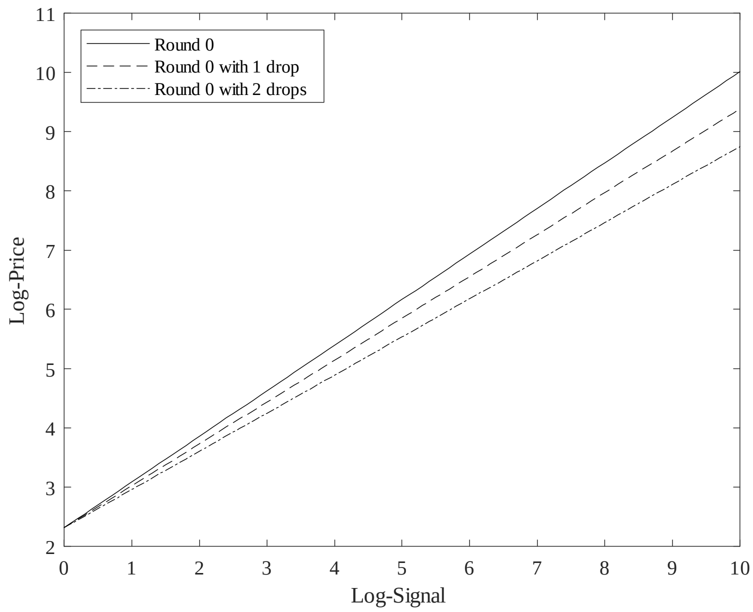
Notes: This graph displays the equilibrium bid functions for $\lambda = 1$ (risk neutrality) and $\lambda = 2.25$ (loss aversion), with bidders bidding substantially lower under loss aversion.

Figure 1.3: Log-Bid Functions in Multiple Rounds



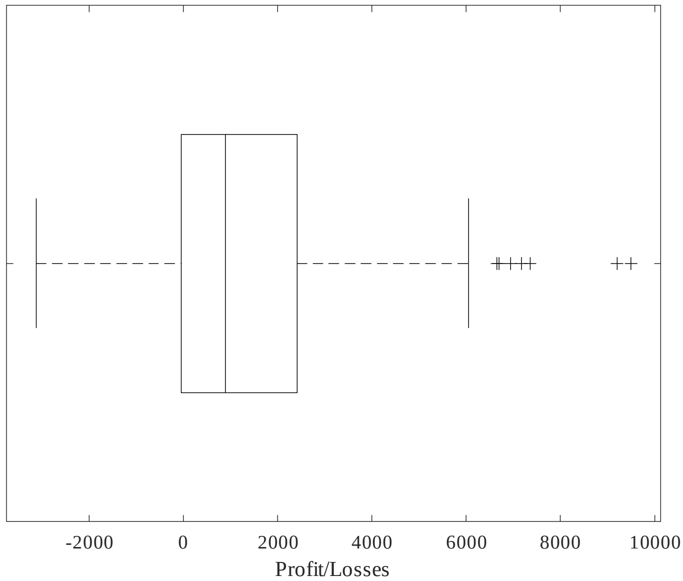
Notes: This graph displays the log-bid functions of a representative bidder for each round in an auction with 5 loss averse bidders.

Figure 1.4: Log-Bid Functions with Active Non-Bidding Participants



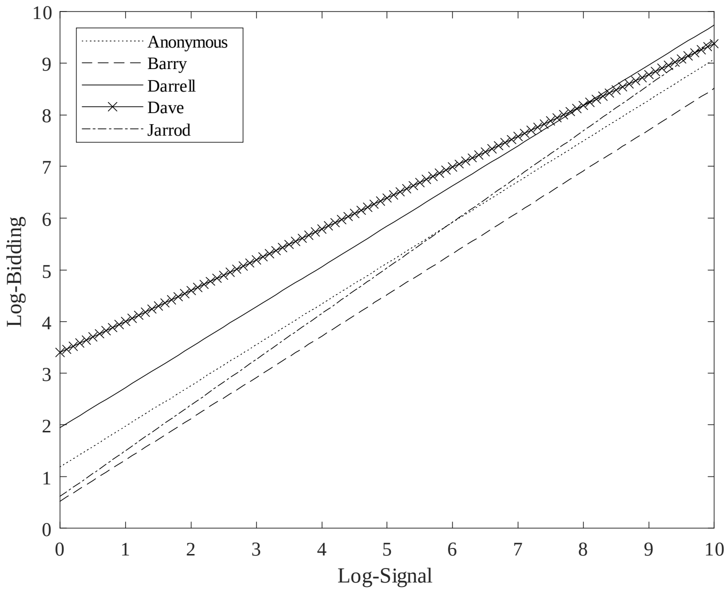
Notes: This graph displays the log-bid function of a representative loss averse bidder when he takes into account the private information active non-bidding participants have in round 0.

Figure 1.5: Distribution of Storage Wars Profit/Losses



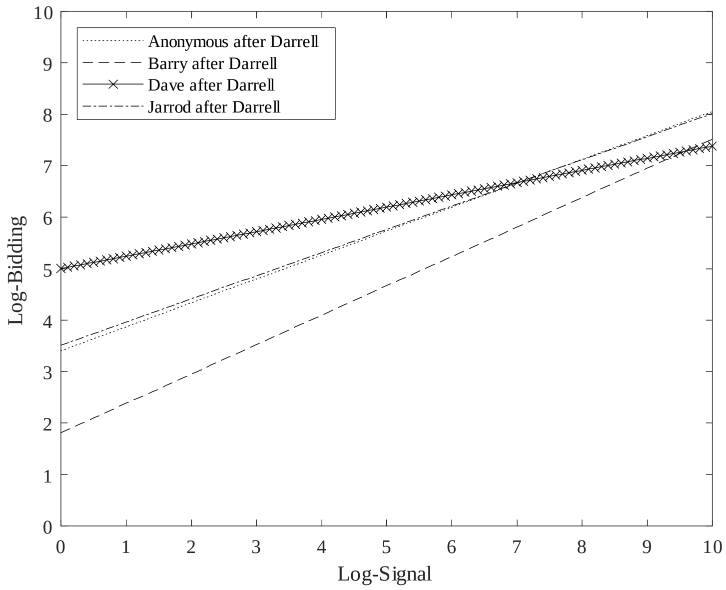
Notes: This graph displays the boxplot of the profit/losses in storage locker auctions, without a few extreme outliers.

Figure 1.6: Log-Bid Functions of Storage Wars Bidders Baseline Model



Notes: This graph displays the round 0 log-bid functions of *Storage Wars* bidders under loss aversion, except for Dave who is risk neutral.

Figure 1.7: Log-Bid Functions of Storage Wars Bidders With Active Non-Bidding Participants



Notes: This graph displays the log-bid functions of *Storage Wars* bidders in round 0 with 1 active non-bidding participant, which in this case is Darrell.

Chapter 2

Heterogeneous Pairs Play Mixed Strategies in the Soccer Field

2.1 Introduction

Mixed strategies are a fundamental component of game theory that allows us to theoretically understand strategic situations which involve unpredictability and mutual outguessing. However, the empirical evidence is still mixed when it comes to assessing whether players actually play consistently with equilibrium predictions.

Experimental situations provide a controlled environment to assess the behavior of players, but they are sometimes criticized because labs might be too aseptic and detached from a real life situation. In contrast, behavior in the field is more likely to reflect real life because of its natural setting, which might provide higher external validity.

The main purpose of this chapter is to check if individuals, when repeatedly facing the same opponents, really behave as game theory predicts by using data from a natural example of strategic play: soccer penalty kicks. A penalty kick can be regarded as a zero-sum game between two players, one kicker and one goalkeeper, because the rules of the game forbid any other player to intervene. They provide a notable advantage over many other real life situations because soccer players are experts at their game and the outcome (goal or miss) is immediately observed after the players choose their strategies.

To test the main implications of mixed strategy equilibrium within pairs, I conducted a (quasi) field experiment in the training grounds of AD Alcorcón, a team from the Spanish Second Division League (also known as LaLiga SmartBank). The dataset I collected includes very detailed information on all relevant aspects of the penalty kicks, specifically the choices taken and the outcome of the kick. The players in my dataset take part in regular competitive leagues in amateur divisions.[1]

Given that in real life situations the same pair of players is rarely observed, previous empir-

[1] Apart from training several hours every day of the working week and playing matches every weekend, they devote a significant fraction of their time and effort to become professional experts in their field.

ical papers using penalty kicks assumed homogeneity of opponents (see Chiappori et al (2002) and Palacios-Huerta (2003, 2017) for examples). In this chapter, I study the consequences of ignoring heterogeneity in empirical work, which arises when pooling observations across rivals. Specifically, I show that assuming homogeneity might lead to false rejections when the different rivals of a given player behave differently. Apart from providing necessary and sufficient conditions for this problem to be irrelevant, I suggest a simple way of combining the test statistics of a player across opponents to obtain a valid aggregate test without making any additional assumptions.

The first testable implication I check is whether the scoring probabilities of a player are identical across strategies, as the theory states they should be. To the best of my knowledge, this is the first time that this hypothesis is tested in the field using repeated observations on specific pairs of kickers and goalkeepers. Empirically, I cannot reject the equality of the scoring probabilities, except for the kickers from the least professional team.

The second testable implication I check is that the actions of the player at each penalty kick should be serially independent, because equilibrium play also requires that each player's choices are independent draws from an *i.i.d.* process. Once again, an important advantage of my dataset is that there are repeated observations for the same pair of players. As in the existing literature, I find that the behavior of most players is consistent with the theory.

The main objective of the next two hypotheses I test is to detect possible interactions between players because a standard assumption in non-cooperative game theory is that players' actions are independent. Specifically, the third hypothesis I check is whether there exists dependence between the strategies of the two players within a given pair at each penalty kick. This hypothesis was already tested by Chiappori et al (2002), who did not reject the null, but they had to pool observations across different, possibly heterogeneous, players because they did not have repeated observations on pairs of kickers and goalkeepers. In contrast, I find dependence between kickers' and goalkeepers' actions for most pairs. The validity of this hypothesis is very important in practice since teams would like to sign goalkeepers that have positive correlation with strikers because it would mean that they can sometimes anticipate where the kicker is going to shoot. At the same time, teams would like to sign strikers who had negative correlation with goalkeepers because it would imply that they are able to deceive them.

In addition, given that in my (quasi) experiment the players participated in a penalty shootout in the training grounds, which is a sequence of penalty kicks in which both kickers and goalkeepers take turns, I also test whether the strategy chosen by consecutive kickers/goalkeepers within teams is influenced by the previous player's strategy. One of the possible alternatives to this hypothesis is that there could exist herd behavior. For example, if the previous kicker shoots to the left, then the next kicker might also decide to shoot to the left, and so on.[2] In this respect, I observe that the goalkeepers of the least professional team tend to replicate each other's actions.

Finally, the availability of repeated observations for each pair of kickers and goalkeepers also allows me to check whether players exhibit some form of learning in the training grounds. In particular, I assess the reinforcement learning model of Erev and Roth (1998), whose main implication is that players respond to negative or positive stimuli by using actions that have worked well for them in the past. However, I find that players do not seem to follow such a reinforcement learning model.

Given that my work contributes to the empirical literature on strategic interactions in two person zero-sum games, I will briefly survey next the existing evidence in professional sports.

Walker and Wooders (2001) tested whether professional tennis players played according to mixed strategies when serving and receiving. Unfortunately, their dataset only contained the server's action and the winner of the point. Still, they found that their data was consistent with the implication of equal payoffs across actions. However, they found negative serial correlation between the actions of a player, i.e. switched actions to often. In contrast, Hsu et al (2007) found that tennis players played consistently with the two implications of the theory using a broader dataset, which included men's, women's, and juniors' matches.

Chiappori et al (2002) offered evidence on the application of mixed strategies to penalty kicks in soccer by testing whether the strategy chosen by the rival forecasts the other player's action in the penalty kick. However, they found no relationship between the kicker's and goalkeeper's actions. Additionally, they could not reject the null hypothesis of equal winning probabilities for the players in their sample. They also tested if there is serial correlation in actions, but they found none. Although their paper represents one of the first attempts to test mixed strategy behavior using data from real soccer games, the nature of their data meant that they looked at

[2] In fact, this type of behavior was observed during the penalty shootout of the 2016 Champions League Final between Real Madrid and Atletico de Madrid, where the kickers from both teams exactly replicated the action of the previous kicker.

the behavior of players aggregating across multiple, possible heterogeneous players rather than at the level of specific kicker-goalkeeper pairs.

One of the contributions of Palacios-Huerta (2002) was to compile a larger dataset, which allowed him to observe repeatedly many individual players. However, he again aggregated across potentially heterogenous opponents because he had little data on repeated interactions of specific pairs. As Chiappori et al (2002), he found that winning probabilities were identical across strategies and that choices were serially independent.

In subsequent work, Palacios-Huerta and Volij (2008) used a 2×2 laboratory experiment borrowed from O'Neill (1987) which mimics penalty kicks. In their simplified lab game, a "kicker" and a "goalkeeper" choose between two actions simultaneously several times. They found that in their lab games professional soccer players played consistently with the mixed strategy equilibrium predictions, with some modest deviations and some serial correlation. In contrast, the rest of the participants did not. However, Levitt et al (2010) found that professional poker, bridge and American soccer players were not able to transfer their professional skills acquired in the field to the lab because they did not behave consistently with equilibrium predictions. This chapter tries to shed light on this conflicting evidence.

The rest of the chapter is organized as follows. Section 2.2 discusses the theoretical setting and its equilibrium. In section 2.3, I discuss the problems that arise from pooling observations of heterogeneous pairs of players as well as suggesting a valid aggregate test. Next, in section 2.4 I describe the dataset that I compiled to test the equilibrium predictions of mixed strategies in the training grounds. The results of the empirical analysis are presented in section 2.5. Finally, section 2.6 studies whether there is evidence of learning. This is followed by the conclusions and several appendices where proofs and additional details can be found.

2.2 Penalty Kicks in Football and Game Theory

2.2.1 *The rules*

According to *Federation Internationale de Football Association* (FIFA) in the Official Laws of the Game (FIFA, 2018) "in soccer, a penalty kick is awarded against a team which commits one of the ten punishable offenses inside its own penalty area while the ball is in play".[3] Additionally, there are penalty shootouts (mostly used in knockout tournaments), which are

[3] The ball is placed on the penalty mark. The goalkeeper remains on his goal line facing the kicker between the goalposts until the ball has been kicked. The rest of the players are located outside the penalty area and they cannot interfere in the kick.

used for determining the winning team in a match that cannot end in a draw after both the regulation and extra playing time have expired.[4]

In a given penalty kick, the ball takes approximately 0.3 seconds to travel the distance between the penalty mark and the goal line. Thus, if the goalkeeper decides his action after the kick, he will not be able to stop the shot (unless, of course, it is aimed at him), so in theory both players must choose their strategies simultaneously.[5] This is one of the hypothesis that I will test in section 2.5.

2.2.2 *The formal setting*

A formal setting of the penalty kick game can be written as follows: one goalkeeper and one kicker are facing each other at a penalty kick. The kicker preferences are to score while the goalkeeper has the opposite preferences, as in all strictly competitive games. Specifically, the kicker's payoff is the probability of scoring while the goalkeeper's payoff is the complementary probability. The kicker may choose to kick to his right (R), to his left (L), or to the center (C). Similarly, the goalkeeper may choose to jump to his left, to his right or remain at the center. When both players choose the same side (L, C, or R) the outcome is less likely to be a goal. In addition, there is usually a natural side for a kicker to shoot, so that the probability of scoring, when kicking to that side, is higher than when kicking to the center or the opposite side, both when the goalkeeper guesses it and when he does not. In contrast, goalkeepers do not have a natural side, but their ability to stop the goal may vary widely across opponents, as I document in detail in section 2.5.

As an example, suppose that the player is right-footed and shooting to his left is his natural side. The payoff matrix, which consists of scoring probabilities, is then:

[4] Each team take turns to shoot five penalty kicks, which must be taken by different kickers. The winning team is decided on the best of five kicks basis. However, if both teams are tied in the number of scored penalties after these five penalty kicks, then the shootout progresses into additional "sudden death" rounds (see Apesteguia and Palacios-Huerta (2010) for more details).

[5] Miller (1998) reports evidence on the fact that both players must choose their strategies simultaneously using data from all the penalty kicks in four World Cups.

		Goalkeeper		
		Left	Center	Right
	Left	$a, 1-a$	$b, 1-b$	$b, 1-b$
Kicker	Center	$e, 1-e$	$c, 1-c$	$e, 1-e$
	Right	$e, 1-e$	$e, 1-e$	$d, 1-d$

where the first payoff corresponds to the kicker and the second payoff to the goalkeeper.

In terms of the payoff matrix, a is the probability that a goal is scored when both players choose the kicker's natural side. But if the kicker is the only one who chooses it, a goal is scored with probability b. Therefore, it makes sense to assume that $b > a$. For any action other than the kicker's natural side, e is the probability that a goal is scored when the actions of both players differ. Similarly, if both players choose action C, a goal is scored with probability c, but if they both choose R, the probability is d. Given that the goalkeeper is more likely to save if he remains at the center, it makes sense to assume that $d > c$. Additionally, as the kicker has a natural side when kicking, then $a > d$. It also makes sense to assume that $b > e$ because the kicker's probability of scoring a goal when kicking to his natural side (in this case left) is higher than when kicking anywhere else, regardless of the actions of the goalkeeper. Finally, it is also reasonable to expect that $e > a$, which means that the kicker is more likely to score when the actions of both players differ.

Under these reasonable conditions, namely $b > e > a > d > c$, there is no pure strategy Nash equilibrium in this game (see section 2.8.2 for a proof). However, there exists a unique mixed strategy Nash equilibrium involving all three strategies where the kicker will choose L, C and R with probabilities

$$p_L = \frac{(e-d)(e-c)}{\Delta}, \; p_C = \frac{(e-d)(b-a)}{\Delta} \text{ and } p_R = \frac{(b-a)(e-c)}{\Delta},$$

where $\Delta = (b+e-a-c)(e-d) - (e-c)(a-b)$. In turn, the goalkeeper will choose the same actions with probabilities

$$q_L = \frac{(e-d)(b-c)+(c-e)(e-b)}{\Delta}, \; q_C = \frac{(e-a)(e-d)}{\Delta} \text{ and } q_R = \frac{(e-a)(e-c)}{\Delta},$$

respectively (see section 2.8.3 for more details). These values guarantee that in expected terms, the probability of the kicker scoring a goal is the same regardless of the strategy chosen (L, C

or R). The same reasoning applies to the goalkeeper.[6]

Given that in my dataset each kicker-goalkeeper pair played the penalty kick game several times, these multiple observations imply that the game was in fact a finite two-person repeated zero-sum game. However, the only subgame perfect equilibrium stipulates to play the Nash equilibrium obtained above in every period (see Osborne (2003) chapter 14 for more details).

From an empirical point of view, the tests would be much simpler if all pairs were alike because one could pool all the observations together. However, this is not the case in practice because the parameters a, b, c, d and e of the payoff matrix depend on the relative abilities of the kicker and goalkeeper, so in general, there will exist pair-specific heterogeneity in the strategies played. For that reason, it is convenient to have repeated observations for each pair. I discuss this issue in more detail next.

2.3 Heterogenous Opponents

The problem of heterogeneity arises when the observations of two or more different pairs of players, each having different abilities or characteristics, are treated as if they all came from the same pair. This is done very frequently in empirical work because of the lack of repeated observations for specific pairs. For example, despite its size, Palacios-Huerta's (2003, 2017) dataset on penalty kicks in actual soccer matches has very few repeated observations for the same pair of kicker-goalkeeper.

In this section, I study the consequences of assuming that the sample observations come from a homogeneous population when in fact it is heterogeneous. For simplicity of exposition, I consider a version of the model in section 2.2.2 with only two actions, although the problem applies more generally. The payoff matrix of this simplified game is the following:

		Goalkeeper	
		Left	Right
Kicker	Left	$a, 1-a$	$b, 1-b$
	Right	$e, 1-e$	$d, 1-d$

where the first payoff corresponds to the kicker and the second payoff to the goalkeeper. As before, the parameters a, b, d and e depend on the relative abilities of the kicker and goalkeeper,

[6] Chiappori et al (2002) discussed a simplified 3×3 payoff matrix which can sometimes give rise to an equilibrium with only two strategies.

so in general, there will exist pair-specific heterogeneity in the strategies played. It is easy to see that there is no pure Nash equilibrium in this game. In this case, the unique mixed strategy Nash equilibrium is such that the kicker will choose L and R with probabilities

$$p_L = \frac{e-d}{b+e-a-d} \text{ and } p_R = \frac{b-a}{b+e-a-d}, \tag{2.1}$$

while the goalkeeper will choose them with probabilities

$$q_L = \frac{b-d}{b+e-a-d} \text{ and } q_R = \frac{e-a}{b+e-a-d}, \tag{2.2}$$

respectively.[7] Again, if the same pair of kicker and goalkeeper play the game repeatedly a finite number of times, the only subgame perfect equilibrium stipulates to play the same Nash equilibrium in every period.

Suppose now the same player faces two different opponents, which gives rise to different parameters values a, b, d and e for each pair, and therefore different equilibrium values. As an illustration, suppose that the payoff matrices for pairs A and B are the following:

Pair A

		Goalkeeper	
		Left	Right
Kicker	Left	0.03,0.97	0.98,0.02
	Right	0.99,0.01	0.02,0.98

Pair B

		Goalkeeper	
		Left	Right
Kicker	Left	0.73,0.27	0.93,0.07
	Right	0.92,0.08	0.89,0.11

where I have chosen these values in such a way that if I choose 20% of the observations from pair A and 80% from pair B, then the average payoff matrix corresponds to the one in Palacios-Huerta (2017). For each of these two pairs, there exists a unique mixed strategy Nash equilibrium

[7] A special case arises when $a = d = 0$ and $b = e = 1$, which leads to a mixed strategy Nash equilibrium where both the kicker and the goalkeeper choose L and R with probability 1/2.

in which the kicker and goalkeeper of pair A play L and R with probabilities $p_L^A = 0.51$ and $p_R^A = 0.49$, and $q_L^A = 0.50$ and $q_R^A = 0.50$, respectively; while the kicker and goalkeeper of pair B will choose L and R with probabilities $p_L^B = 0.13$ and $p_R^B = 0.87$, and $q_L^B = 0.17$ and $q_R^B = 0.83$, respectively.

The joint probability distribution in the population for the kicker-goalkeeper pair i, for $i = A, B$, is

Direction/Outcome	*Left*	*Right*	*Sum*
Success	π_{SL}^i	π_{SR}^i	π_S^i
Failure	π_{FL}^i	π_{FR}^i	π_F^i
Sum	π_L^i	π_R^i	1

where π_L^i denotes the marginal probability of the left strategy and π_S^i the marginal probability of scoring.

Suppose now that an empirical researcher erroneously treats all the observations as though they came from the same pair. Let's define a heterogeneous player H as drawn with probability λ from pair A and with probability $1-\lambda$ from B. The following proposition establishes necessary and sufficient conditions that allow researchers to ignore heterogeneity (see section 2.8.1 for proof).

Proposition 2.1 *If both pairs of kicker and goalkeeper play consistently with the equilibrium predictions of mixed strategies, the scoring probabilities of a "heterogeneous" player will differ across strategies unless $\lambda = 0$ or $\lambda = 1$ (no heterogeneity in the sample), $\pi_L^A = \pi_L^B$ (no heterogeneity in the strategies) or $\pi_S^A = \pi_S^B$ (no heterogeneity in the outcome).*

As a special case, this proposition justifies the sufficient condition of identical goalkeepers in Chiappori et al (2002), who stated that if goalkeepers are indeed homogeneous, the kicker's strategy will be independent of the goalkeeper he is facing. In contrast, the result in Proposition 2.1 is both necessary and sufficient. In this sense, it is important to emphasize that it is not enough that one of the players is the same across pairs; even in that case, what matters is whether $\pi_L^A = \pi_L^B$ or $\pi_S^A = \pi_S^B$.

To investigate the effects of heterogeneity, I study the correlation between the actions of a supposedly homogeneous player with his scoring probability, which should be equal to zero under the null $H_0 : \pi_{hj}^i = \pi_h^i \times \pi_j^i$, for $h = S, F$ and $j = L, R$, where S (success), F (failure), L (left) and R (right) are dummy variables with $F = 1 - S$ and $R = 1 - L$.

If both pairs of kicker and goalkeeper play consistently with equilibrium predictions, then the marginal probability of their actions (π_L^i) is equal to the mixed strategy equilibrium (p_L^i, q_L^i) for each pair i in (2.1) and (2.2). Additionally, given independent actions between players, the marginal probability of scoring can be easily computed as

$$\pi_h^i = \sum_{j=L}^{R} \Pr(h=1|j,m) * p_j^i * q_m^i, \text{ where } m = L, R.$$

However, the marginal probability of scoring and the marginal probability of the actions of the heterogeneous "player" H will be $\pi_j^H = \lambda\pi_j^A + (1-\lambda)\pi_j^B$ and $\pi_h^H = \lambda\pi_h^A + (1-\lambda)\pi_h^B$, respectively. Therefore,

$$Corr(h^H, j^H) = \rho_H = \frac{Cov(h^H, j^H)}{\sqrt{Var(h^H)Var(j^H)}}, \tag{2.3}$$

where

$$\begin{aligned} Cov(h^H, j^H) &= \lambda(1-\lambda)\left(\pi_h^A - \pi_h^B\right)\left(\pi_j^A - \pi_j^B\right) \\ Var(h^H) &= \pi_h^H(1-\pi_h^H) \\ Var(j^H) &= \pi_j^H(1-\pi_j^H) \end{aligned}$$

(see section 2.8.1 for more details).

(Figure 2.1)

Figure 2.1 shows the effects of varying the fraction of observations from pair A (λ) on the correlation between the actions of the supposedly homogeneous player with his scoring probability. As expected, $\lambda = 0$ or $\lambda = 1$ (no heterogeneity in the sample) implies that player H is indeed homogeneous. However, for any other value of λ, there is an apparent dependence between a player's actions and his scoring probabilities when in fact there is none.

(Figure 2.2)

Figure 2.2 shows the theoretical rejection rates obtained with Pearson's independence test statistic with 2 actions and 1 degree of freedom for 20, 38 (Palacios-Huerta (2003) median sample size) and 200 observations as a function of λ. As can be seen, when $\lambda = 0$ or $\lambda = 1$, the rejection rate is the nominal (5%) size, while for any other value of λ, it exceeds 5%. This occurs even though both underlying pairs of players play consistently with the theory. Therefore, the homogeneity assumption misleadingly increases the rejection rate of the test statistics because the non-centrality parameter of the distribution of the statistic is $n\rho_H^2$, where

ρ_H is the correlation between the actions of the supposedly homogeneous player with his scoring probability defined in (2.3). As is well known, the distribution of a non-central $\mathcal{X}^2$ shifts to the right as the non-centrality parameter $n\rho_H^2$ increases, which implies an increase in the rejection rate (see Mood et al (1974) for more details). Additionally, in section 2.8.4 I confirm that the same conclusions hold in Monte Carlo simulations. However, it is important to note that the rejection rate is rather low unless the sample size is large (see Figure 2.2 for more details).

So in summary, under heterogeneity, researchers may mistakenly reject the null when testing the implications of mixed strategy even though the null is true.

2.3.1 *Allowing for heterogeneity of opponents*

I propose a simple solution to the heterogeneity problem in those situations in which there are multiple observations for all the pairs involving a given player. The intuition is as follows. For a given pair of players, all the independence test statistics converge to a $\mathcal{X}^2$ under the null when the number of observations goes to infinity.[8] Therefore, I can compute an aggregate test for a given player as the sum of the independent $\mathcal{X}^2$ statistics across all his opponents, which results in another $\mathcal{X}^2$ with degrees of freedom equal to the sum of degrees of freedom for each pair test.

Specifically, suppose that the same kicker plays against N different goalkeepers. Algebraically, his aggregate $\mathcal{X}^2$ test statistic will be:

$$\Psi = \sum_{i=1}^{N} \Psi_i,$$

where Ψ_i, for $i = 1, \ldots, N$, is his $\mathcal{X}^2$ independence test statistic obtained from the observations he shares with his i^{th} opponent.

From an empirical point of view, this simple aggregate test statistic allows both the optimal mixed strategies and the scoring probabilities to be different for different opponents. In addition, it allows the theoretical results to be tested player by player, thereby using more observations for each player than each specific pair test. Obviously, if there is a single observation per pair, this procedure cannot be applied. But as the number of observations per pair increases, its reliability will increase.

Heterogeneity also affects the tests of serial correlation and action independence, but the

[8] As is well known, those tests which converge to an F distribution with ν_1 and ν_2 degrees of freedom can be converted into $\chi^2_{v_1}$ by multiplying the F statistic by ν_1.

problems are very similar, so I will not discuss them separately.

2.4 (Quasi) Field Experiment

2.4.1 *Soccer subjects*

As I explained in the introduction, I conducted a (quasi) field experiment in the training grounds of AD Alcorcón, a team from the Spanish Second Division League (also known as LaLiga SmartBank), for periods of 15 minutes each day over a three week period in April 2016. The players came from AD Alcorcón youth teams, which were taking part in regular league competitions in amateur divisions. Those leagues have the same structure, calendar schedule and rules as professional leagues (FIFA, 2018).

There are two types of players, who differ in seniority: "Cadetes" and "Juveniles". The players from the Cadete teams are 15 and 16 years old and the players from the Juvenil team 17, 18 and 19 years old. AD Alcorcón has three Juveniles teams. The players I recruited from that category come from the Juvenil A team which plays in the Honor Division, the top level of the Spanish soccer league system for youth players. Those players are in their last formative stages and aspire to climb the last step that leads them to Alcorcón B, the reserve team of the first team. Moreover, there are three Cadete teams. The players I recruited from this age group play in the Primera Division Autonomica and Preferente Cadete, which are the highest and second highest divisions in that category, respectively.

2.4.2 *Experimental setup*

Once all the players were recruited, I was able to create several pairs of kickers and goalkeepers within each team. Since there are more defenders, midfielders and forwards than goalkeepers, the latter were paired at least four times. In particular, every goalkeeper was matched with one left and right-footed penalty kick specialist with substantial experience in kicking penalties, as well as with left and right-footed inexperienced penalty kickers.[9] There are a total of 14 kickers and 6 goalkeepers in the dataset.[10]

After the pairs of kickers and goalkeepers were formed, they played a penalty shootout in the training grounds, which is simply a sequence of penalty kicks where players take turns, with a random initial order. This was done so as to have approximately the same number

[9] The selection was made by the managers of the different teams who knew the players' abilities well. This selection should in principle increase the heterogeneity within pairs.

[10] Due to the confidentiality agreements I signed with the players' agents, I cannot reveal any personal identifying information.

of observations for each pair. Both kickers and goalkeepers regularly alternated to maintain a high level of concentration in each practice session. Given that penalty kicks decide matches, qualifications for next rounds in tournaments and even titles, soccer teams devote considerable resources to analyze and improve strategies for their players. For that reason, coaches told players that the penalty shootouts in my (quasi) experiment were an integral part of their training.

2.4.3 *Descriptive statistics*

There is a total of 8 pairs from Cadete A with 16 penalty kicks each on average. Moreover, in Cadete C, there are also 8 pairs with approximately 13 penalties each. In Juvenil A, there are 10 pairs with 10 penalties each on average. For each of these pairs, the observations in the dataset include all the penalties they participated, in chronological order. Given that the different teams played the penalty kick game over non-consecutive days, I have taken these breaks into account in some of the tests.

The dataset includes the date and time at which the penalty kick took place, the identifying codes of the kicker and goalkeeper for each penalty kick, the choices taken (L, C and R), the foot used by the kicker (left or right), and the outcome of the kick (goal or miss). There were two independent measurements taken for each penalty kick to eliminate measurement error.[11]

Table 2.1 offers a basic description of the data. It shows the relative proportions of choices made by both kickers and goalkeepers from the different AD Alcorcón youth teams (L, C or R). The first letter refers to the choice made by the kicker and the second one to the choice made by the goalkeeper, always from the point of view of the goalkeeper. For instance, R-L means that the kicker chooses to kick to the right hand side of the goalkeeper (the natural side of a right-footed player) and the goalkeeper chooses to jump to his left. The last column shows the scoring rate for a given team.

(Table 2.1)

The strategy followed by goalkeepers coincides with that followed by kickers in 47.84% of all penalties in the dataset. Kickers do not usually kick to the center (11.71% of all kicks), whereas goalkeepers remain in the middle less often (8.78%). The percentage of kicks where the actions of the players do not coincide is mostly divided between L-R (18.76%) and R-L (19.35%). A

[11] One measurement was taken from the point of view of the kickers while the other one from the goalkeepers.

goal is scored in 75.66% of all penalty kicks. The scoring rate is over 88.84% when the kicker choice differs from the goalkeeper, but it is just over 51.41% when it coincides.

Moreover, the scoring rate of penalties of Juvenil A, Cadete A and Cadete C are 76.92%, 72.86% and 77.77%, respectively. It may seem surprising that the scoring rate of the Cadete C, which has the least professional players in the sample, is the highest of all the teams. This is because the goalkeepers saving rate from Cadete C team is under 40% in all of their strategies, the worst of the three teams (see section 2.5.1.1 for more details).

2.5 Empirical Analysis

2.5.1 *Test of equal scoring probabilities*

The first testable implication I check is whether the scoring probabilities for a player are identical across strategies. Following the discussion in sections 2.8.5 and 2.8.6 regarding the size and power of the different tests proposed in the literature, I use the F-test version of the Linear Probability Model (LPM) (see Wooldridge (2002) chapter 7 for more details).[12]

Let S take the value 1 if the penalty is scored and 0 otherwise. Given that each player (kicker/goalkeeper) has three strategies available (Left "L", Center "C", and Right "R"), the LPM can be written as:

$$S = \delta_L L + \delta_C C + \delta_R R + u, \tag{2.4}$$

where L, C and R are mutually exclusive dummy variables and u has zero conditional mean, i.e. $E(u|L, C, R) = 0$. For example, L takes the value 1 if the penalty is shot in that direction and 0 otherwise.

The regression coefficients of the LPM have a direct interpretation as conditional scoring probabilities. For instance, δ_L is the proportion of left kicks scored. Thus, the estimated probabilities are always non-negative and they add up to 1, which avoids a common criticism of the LPM (see again Wooldridge (2002)).

The null hypothesis of equal scoring probabilities states that $\delta_L = \delta_R = \delta_C = \delta$. In practice, it is easier to test this hypothesis by estimating the following modification of model (2.4):

$$S = \beta_0 + \beta_1 L + \beta_2 R + u, \tag{2.5}$$

[12] There are many econometric procedures to test this hypothesis. However, Proposition 3.1 of chapter 3 proves the numerical equivalence between Pearson's contingency table test for independence and the Lagrange Multiplier (LM) and overidentifying restrictions test in several popular linear and non-linear regression models. Therefore, the results that I will present are largely insensitive to the methodology used.

where $\beta_0 = \delta_C$, $\beta_1 = \delta_L - \delta_C$ and $\beta_2 = \delta_R - \delta_C$. In (2.5), the coefficients of the dummy variables are the differences between the scoring probabilities of the corresponding strategy and the baseline, which in this case is C. The fit of this beta regression is identical to the fit of the regression in (2.4), but it has the advantage that the null hypothesis of equal scoring probabilities can be expressed as $\beta_1 = \beta_2 = 0$. This can be tested using an F-test with 2 degrees of freedom in the numerator and $n-3$ degrees of freedom in the denominator, where n is the number of observations.[13] The exact formula of this F-test is:

$$F = \frac{(R^2/k)}{(1-R^2)/(n-k-1)},$$

where R^2 measures the proportion of the variability of the dependent variable explained by the k non-constant explanatory variables. Therefore, the F-statistic would be 0 if the scoring probability is exactly the same across strategies (single outcome) and/or if the player is only employing one strategy (single choice). However, the F-statistic would be infinity when the regressors provide a perfect fit, i.e. $R^2 = 1$ (see section 2.8.5 for more details).

The LPM has one potentially important disadvantage. Under the alternative, it violates the homoskedasticity assumption because the conditional variance of the error term u will change depending on the values of the explanatory variables (see Wooldridge (2002)). However, the variance of u given the dummy regressors is constant under the null hypothesis of equal scoring probabilities ($\beta_0(1-\beta_0)$). This implies that the homoskedasticity assumption holds and the F-test is valid.

2.5.1.1 Pair tests

As I mentioned before, an important advantage of my dataset is that for the first time I have repeated observations for each and every pair of kickers and goalkeepers. Therefore, I can carry out separate tests that check whether each member of the pair within a team is playing consistently with equilibrium outcomes. There is a total of 26 pairs in the dataset, and for each pair, there is a test statistic for the kicker and another one for the goalkeeper. However, it is important to note that the pair tests alone may have low power because of the relative low number of observations (see section 2.8.6 for more details). The results of all those tests are shown in Tables 2.2, 2.3 and 2.4. Table 2.2 corresponds to Cadete A, Table 2.3 to Cadete C and

[13] Some of the players in the training grounds never employed one of the three strategies (either L, C or R). When that occurs, the F-test will have 1 degree of freedom in the numerator and $n-2$ degrees of freedom in the denominator.

Table 2.4 to Juvenil A.

(Table 2.2)

(Table 2.3)

(Table 2.4)

The rejections I find only come from the kickers and goalkeepers from the Cadete A and Cadete C team. Therefore, my evidence is trivially consistent with the first implication for all the players from the Juvenil A team.

It is worth mentioning that the scoring rates vary substantially across pairs of the three teams, which confirms the empirical relevance of the discussion in section 2.3. An interesting observation I found was that kicker 3 of Juvenil A, who is a left-footed penalty kick specialist, had a test statistic of 0 in all of the pair tests. This is because his scoring probabilities are 100% regardless of the strategy chosen, so not only is he a great performer, but he also behaves perfectly according to the theory.

2.5.1.2 Tests allowing for heterogeneity of the opponents

Given that the pair tests do not exploit the fact that a player is matched several times, I compute an additional test that checks whether each player behaves as the theory predicts when aggregating all his observations but without assuming homogeneity of his opponents. This test should have substantially more power than each specific pair test. The results are shown in Table 2.5.

(Table 2.5)

Panel A describes the results for individual players for the Cadete A team. The null hypothesis is rejected for one kicker and one goalkeeper at the 5% significance level and one additional kicker at the 10% level. Panel B shows the results for the Cadete C team. The null hypothesis is rejected for two kickers at the 5% significance level and one goalkeeper at the 1% level. Finally, Panel C includes the results for the Juvenil A team. The hypothesis is only rejected for one kicker at the 10% level.

Hence, the evidence obtained by aggregating each player's opponents is consistent with the first implication for the goalkeepers from the Juvenil A team. As for the kickers from the Cadete A team, if I take into account that there are multiple tests (see section 2.8.7), the binomial probability of one or more kickers out of 4 rejecting the null at the 5% level when the null is true is 0.185, so the evidence suggests that as a group, those players do not reject the null hypothesis either. Additionally, the evidence I find for the kickers from the Juvenil A team is also consistent with the theory because the probability of one or more kickers out of 6 rejecting the null at the 10% level is 0.468.

However, given that the probability that two or more kickers out of 4 rejecting the null at the 5% level is 0.014, I can claim that the scoring probabilities of the kickers from the Cadete C team differ depending on the action. As for the goalkeepers from the Cadete A and Cadete C teams, the evidence is more mixed because the probability of at least one goalkeeper out of 2 rejecting the null at the 5% and 1% level is 0.097 and 0.02 respectively.

In contrast, I find that the null hypothesis of equal scoring probabilities is rejected for the kickers from the Cadete A team when I incorrectly treat all their opponents as if they were a single homogeneous one because the probability of two or more kickers out of 4 rejecting the null at the 5% level is 0.014. This false rejection confirms the importance of recognizing the heterogeneity of opponents.

2.5.2 *Test for serial independence*

The second testable implication I check is that the actions of the player at each penalty kick should be serially independent. In that regard, note that the players' strategies will not be serially independent if they switch actions too often (negative serial correlation) or if they choose not to switch their actions regularly (positive serial correlation). Following the discussion in sections 2.8.5 and 2.8.6 regarding the size and power of the different tests proposed in the literature, I use the F-version of the Lawley-Hotelling trace test (LH) in the multivariate version of the LPM to test if the player's strategies are serially independent (see Stewart (1995) for more details).[14]

The multivariate regression I have used to detect possible departures from serial independence is similar to a first-order vector autoregressive process for dummy variables (see Wooldridge

[14] The numerical equivalence results Proposition 3.1 of chapter 3 also applies to tests of serial independence of a discrete Markov chain, which can be regarded as an analog to the multinomial model, although in a time series context. Therefore, the results that I present should be largely insensitive to the methodology.

(2002) chapter 18, section 5 for more details). Specifically,

$$\begin{pmatrix} L_t \\ C_t \\ R_t \end{pmatrix} = \begin{pmatrix} \delta_{LL} & \delta_{CL} & \delta_{RL} \\ \delta_{LC} & \delta_{CC} & \delta_{RC} \\ \delta_{LR} & \delta_{CR} & \delta_{RR} \end{pmatrix} \begin{pmatrix} L_{t-1} \\ C_{t-1} \\ R_{t-1} \end{pmatrix} + \begin{pmatrix} u_{Lt} \\ u_{Ct} \\ u_{Rt} \end{pmatrix},$$

where L_t, C_t and R_t are the dependent variables, L_{t-1}, C_{t-1} and R_{t-1} are lagged regressors, δ_{CL} measures the probability of L_t being equal to 1 given that C_{t-1} is equal to 1, etc. In this multivariate regression with three lagged explanatory variables, but no constant, the coefficients of the lagged variables are the probability of choosing a strategy at time t conditional on the previous action. These are sometimes called transition probabilities. The sum of δ_{LL}, δ_{LC} and δ_{LR} is equal to 1, and the same applies to the other columns in the matrix. Therefore, the coefficients in equation C_t can be obtained from the other two equations because $C_t = 1 - L_t - R_t$. For that reason, I can eliminate this equation from the system of equations without loss of generality to avoid the singularity (see Judge et al (1985) chapter 12, section 5 for more details).

The null hypothesis of serial independence implies that $\delta_{LL} = \delta_{CL} = \delta_{RL}$ and $\delta_{LR} = \delta_{CR} = \delta_{RR}$. In practice, it is easier to test this hypothesis by estimating the following model:

$$\left.\begin{aligned} L_t &= \beta_{L0} + \beta_{LL} L_{t-1} + \beta_{LR} R_{t-1} + u_{Lt} \\ R_t &= \beta_{R0} + \beta_{RL} L_{t-1} + \beta_{RR} R_{t-1} + u_{Rt} \end{aligned}\right\},$$

where $\beta_{L0} = \delta_{CL}$ and $\beta_{R0} = \delta_{CR}$, $\beta_{LL} = \delta_{LL} - \delta_{CL}$, $\beta_{RL} = \delta_{RL} - \delta_{CL}$, $\beta_{LR} = \delta_{LR} - \delta_{CR}$ and $\beta_{RR} = \delta_{RR} - \delta_{CR}$. In the regression with only two lagged variables and a constant, the coefficients of the lagged variables are the differences between the probabilities of the corresponding strategy and the baseline, which is the lagged variable C_{t-1}. The adjustment of these regressions is identical to the adjustment of the regressions written in terms of δ's, but they have the advantage that the null hypothesis of serial independence can be expressed as
$\beta_{LL} = \beta_{LR} = \beta_{RL} = \beta_{RR} = 0$. In addition, homoskedasticity will again hold under the null, so the usual regression tests remains valid.

2.5.2.1 Pair tests

As in section 2.5.1.1, given that I have repeated observations on each pair of kicker-goalkeeper, I can compute a test for the kicker and another one for the goalkeeper to check whether the null hypothesis of serial independence holds using multiple observations for each and every pair. But as I mentioned in section 2.4, the different teams played the penalty shootout over non-consecutive days, so there were long breaks between some of the observations. For that reason, instead of assuming that the players remembered what they did at the very end of the previous day, I test for serial correlation within each practice session, but combine the different sessions for a given pair. In practice, this means dropping the first observation from each day. This allowed me to have a larger sample for each pair, which enables the test to have more power to reject the null. The results of the tests are shown in the following tables:

(Table 2.6)

(Table 2.7)

(Table 2.8)

The null hypothesis of serial independence is only rejected for the kickers in pair 4 and 9 from the Juvenil A team at the 5% and 1% level respectively, which is surprising because they are both two penalty kick specialists. In fact, the actions of pair 9 kicker provide a perfect fit (see section 2.8.5.1 for more details).

2.5.2.2 Tests allowing for heterogeneity of the opponents

Following the discussion in section 2.5.1.2, I also check whether the behavior of each individual player is consistent with this second implication when aggregating all his observations but without assuming homogeneity of his opponents. Again, the solution is to add up the $\mathcal{X}^2$ versions of the pair tests. The results are shown in Table 2.9.

(Table 2.9)

Panel A describes the results for individual players from the Cadete A team while Panel B shows the results for the Cadete C team. The hypothesis of serial independence is not rejected for any of those players, implying that they are indeed able to generate random sequences even though they are not the most professional players in the sample. Finally, Panel C includes the results for the Juvenil A team. The null hypothesis is only rejected for two kickers at the 10% level.

In this context, one could therefore argue that most of the evidence obtained by aggregating each player's opponents is consistent with the second implication for all of the players in the sample, including the kickers in the Juvenil A team because the probability of two or more kickers out of 6 rejecting the null at the 10% level is 0.114 (see again section 2.8.7). Thus, they seem truly able to generate random sequences; they do not appear to switch strategies too regularly or to seldom. This differs from the evidence of negative serial dependence in Walker and Wooders (2001), who tested whether professional tennis players played according to the theory when serving and receiving.

2.5.3 *Test for action independence*

In a penalty kick, both players must choose their strategies simultaneously due to the nature of the game (see the discussion in Miller (1998) and footnote 4, section 2.2.1 for more details). Therefore, an important implication of their randomizing behavior is that there should be no dependence between the strategies played by the two players. Thanks to the repeated nature of my data, I can follow a similar approach as in the previous section to test for possible interactions for each pair of players from the three teams. In fact, the econometric procedure is analogous to the one used for testing serial independence described in section 2.5.2, except that here regressands and regressors correspond to the same time period and the explanatory variables correspond to the actions of his opponent (see section 2.8.6 for more details).[15] Obviously, this test can only be done at the pair level. The results for the three teams are displayed in Tables 2.10, 2.11 and 2.12.

(Table 2.10)

(Table 2.11)

(Table 2.12)

The results show that of the 26 existing pairs, the null hypothesis is rejected for one pair from the Cadete A team, and two pairs from the Cadete C team at the 10% level, one pair from the Juvenil A at the 5% level, and finally two pairs from the Cadete C team and one pair from Juvenil A team at the 1% level. In fact, in the Cadete C team, the regressors for pairs 5 and 6 provide a perfect fit. These two pairs correspond to both goalkeepers playing against the same inexperienced left-footed kicker.

If I take into account that there are multiple tests (see section 2.8.7), the probability that two or more pairs out of 8 from the Cadete C team rejecting the null at the 1% level is 0.002, so there seems to be dependence between the kicker's and goalkeepers actions in a penalty kick. This is

[15] It is worth mentioning that the results in Proposition 3.1 of chapter 3 imply that I would get the same results if I exchanged regressors and regressands in these regressions.

not very surprising because they are the least experienced players in the sample. Similarly, as the probability that at least one pair out of 10 from the Juvenil A team rejecting the null at the 1% level is 0.095, there is marginal evidence of dependence between the kicker's and goalkeeper's actions. In contrast, I can conclude that the kickers' and goalkeepers' actions are not correlated for the Cadete A team players because the probability of one or more pairs out of 8 rejecting the null at the 10% level is 0.569. This differs from the evidence in Chiappori et al (2002), who did not rejected the null, but they did not have repeated observations. However, my finding is in line with the results in Belot et al (2013).

2.5.4 *Test for sequential independence*

Finally, I test that the strategy chosen by consecutive kickers/goalkeepers within teams is independent of the previous player's strategy. Recall that players played a penalty shootout in the training grounds with an initial random order, so that both kickers and goalkeepers regularly alternated to maintain a high level of concentration in each practice session. This hypothesis will be rejected if there is herd behavior. For example, if the previous kicker shoots to the left, then the next kicker might also decide to shoot to the left, and so on. Once again, instead of assuming that the players remembered what they did at the very end of the previous day, I test for sequential independence by combining the different sessions for a given team without the first observation from each day.

The econometric procedure is analogous to the one described in section 2.5.2 except that now the lagged variables represent the action of different kickers/goalkeepers from the same team. The results of the test are shown in Table 2.13.

(Table 2.13)

The main result in this analysis is that the null hypothesis is only rejected for the goalkeepers from the Cadete C team at the 5% significance level. Apparently, the goalkeepers from that team tended to replicate the strategy of the previous goalkeeper. This may occur because these players have less years of experience and play in the least competitive league of the three teams in the sample. On this basis, one could say that those players exhibited some form of herd behavior. In contrast, players with substantial experience tend to rely on their own actions and not on the previous players actions.

2.6 Reinforcement Learning

Although many economic theories rely on the analysis of Nash equilibria in games, they do not necessarily require fully rational players. In fact, Nash equilibrium might arise as a result of less than fully rational players learning over time. For that reason, I study if the reinforcement learning model of Erev and Roth (1998), whose main implication is that players respond to negative or positive stimuli by using actions that have worked well in the past, might be relevant for the players in my dataset.

Assume that at $t = 1$, each player i has an initial propensity to play his m^{th} pure strategy (L, C or R), given by $\vartheta_{im}(1)$. For simplicity, assume that to begin with, each player i will have equal propensities for each of his pure strategies, so $\vartheta_{iL}(1) = \vartheta_{iC}(1) = \vartheta_{iR}(1) = 1/3$. After each play, propensities are updated using a reinforcement function. Specifically, assume that if player i plays his m^{th} pure strategy at time t and obtained a payoff x, then his propensity to play strategy m at time $t + 1$ is updated by setting

$$\vartheta_{ih}(t+1) = \left.\begin{array}{ll} \vartheta_{im}(t) + R(x) & \text{if } m = m_t \\ \vartheta_{im}(t) & \text{otherwise} \end{array}\right\},$$

for some increasing function $R(\cdot)$. The idea is that if m_t was successful, the player is more likely to use that strategy again. However, if it was unsuccessful, he will be less likely to play it.

Propensities are mapped into choices using a probabilistic choice rule. For instance, letting $\upsilon_{im}(t)$ denote the probability that player i will choose action m at time t, a simple rule would be:

$$\upsilon_{im}(t) = \frac{\vartheta_{im}(t)}{\sum_{m=L,C,R} \vartheta_{im}(t)},$$

where the sum is taken over all player i's pure strategies (L, C and R).

Therefore, a testable implication of the reinforcement learning model is that a player's strategy depends on the outcome of his previous action. On the other hand, if kickers and goalkeepers play according to mixed strategies, then they will not modify the probabilities of their actions regardless of the outcome of their previous actions.[16]

To test if players use such a learning mechanism, I will use the multivariate version of the LPM in section 2.5.2 but this time using as regressors interaction terms between the lagged outcome (success (S_{t-1}) and failure ($F_{t-1} = 1 - S_{t-1}$)) and the lagged regressors (L_{t-1}, C_{t-1}

[16] This hypothesis is somewhat related to the implication of serial independence but it is substantially different.

and R_{t-1}). Specifically, I consider

$$\begin{pmatrix} L_t \\ C_t \\ R_t \end{pmatrix} = \begin{pmatrix} \phi_{LLS} & \phi_{CLS} & \phi_{RLS} & \phi_{LLF} & \phi_{CLF} & \phi_{RLF} \\ \phi_{LCS} & \phi_{CCS} & \phi_{RCS} & \phi_{LCF} & \phi_{CCF} & \phi_{RCF} \\ \phi_{LRS} & \phi_{CRS} & \phi_{RRS} & \phi_{LRF} & \phi_{CRF} & \phi_{RRF} \end{pmatrix} \begin{pmatrix} LS_{t-1} \\ CS_{t-1} \\ RS_{t-1} \\ LF_{t-1} \\ CF_{t-1} \\ RF_{t-1} \end{pmatrix} + \begin{pmatrix} u_{Lt} \\ u_{Ct} \\ u_{Rt} \end{pmatrix},$$

where L_t, C_t and R_t are the dependent variables, mh_{t-1}, for $m = L, C, R$ and $h = C, F$, is an interaction term between the lagged regressors and the lagged outcome, and ϕ_{LLS} measures the probability of L_t being equal to 1 given that LS_{t-1} is equal to 1, etc. As usual, the coefficients in equation C_t can be obtained from the other two equations because $C_t = 1 - L_t - R_t$. For that reason, I eliminate this equation from the system of equations without loss of generality to avoid the singularity.

The null hypothesis of no learning implies that $\phi_{LjS} = \phi_{CjS} = \phi_{RjS} = \phi_{LjF} = \phi_{CjF} = \phi_{RjF}$ for $j = L, R$. In practice, it is easier to test this hypothesis by estimating the following model:

$$\left.\begin{aligned} L_t &= \omega_{L0} + \omega_{LLS}LS_{t-1} + \omega_{CLS}CS_{t-1} + \omega_{RLS}RS_{t-1} + \omega_{LLF}LF_{t-1} + \omega_{RLF}RF_{t-1} + u_{Lt} \\ R_t &= \omega_{R0} + \omega_{LRS}LS_{t-1} + \omega_{CRS}CS_{t-1} + \omega_{RRS}RS_{t-1} + \omega_{LRF}LF_{t-1} + \omega_{RRF}RF_{t-1} + u_{Rt} \end{aligned}\right\},$$

where $\omega_{L0} = \phi_{CLF}$, $\omega_{R0} = \phi_{CRF}$, $\omega_{LLS} = \phi_{LLS} - \phi_{CLF}$, $\omega_{LRS} = \phi_{LRS} - \phi_{CRF}$, etc. In these regressions with only five variables and a constant, the coefficients of the lagged explanatory variables are the differences between the probabilities of the corresponding strategy's outcome and the baseline, which corresponds to CF_{t-1}. Otherwise, the econometric procedure is analogous to the one used to test for serial independence described in section 2.5.2.

2.6.1 *Pair tests*

Given that there are multiple observations for each pair of kickers and goalkeepers, I check whether the null hypothesis of lack of learning holds. As usual, for each pair there is a test statistic for the kicker and another one for the goalkeeper. The results of the tests are shown in

Tables 2.14, 2.15 and 2.16.

(Table 2.14)

(Table 2.15)

(Table 2.16)

In the Cadete A team, pair 3 kicker rejects the null hypothesis of lack of learning at the 10% significance level. Additionally, in the Cadete C team, the null is only rejected for pair 3 kicker at the 5% level and for pair 6 goalkeeper at the 10% level. Similarly, in the Juvenil A team, pair 4 kicker and pair 8 goalkeeper reject the null at the 10% level. Despite these rejections, only a few of the players showed clear evidence of reinforced learning. For instance, the kickers from pair 3 from the Cadete A and Cadete C teams seemed to change strategies when they missed and remain playing the same strategy if they scored. In contrast, pair 4 kicker and pair 8 goalkeeper from the Juvenil A team did not play according to the implications of reinforced learning because surprisingly they switched strategies too often whenever at $t-1$ the outcome was a goal.

2.6.2 *Tests allowing for heterogeneity of the opponents*

Once again, I compute an additional test that checks whether each player behaves consistently with learning by aggregating all his observations but without assuming homogeneity of his opponents. The results are shown in Table 2.17.

(Table 2.17)

Panel A describes the results for individual players for the Cadete A team. The null hypothesis of lack of learning is not rejected for any of the players. Panel B shows the results for the Cadete C team. The null hypothesis is rejected for one kicker and one goalkeeper at the 10% level. Finally, Panel C includes the results for the Juvenil A team. The hypothesis is rejected for one kicker at the 10% level.

Nevertheless, if I take into account that there are multiple tests (see section 2.8.7), most of the evidence obtained by aggregating each player's opponents does not suggest the presence of reinforced learning because the probability of one or more kickers out of 4 from the Cadete C team rejecting the null at the 10% is 0.344 while the probability of one or more kickers out of 6 from the Juvenil A team rejecting the null at the 10% level is 0.468.

2.7 Conclusions

In this chapter I conducted a (quasi) field experiment in the training grounds of AD Alcorcón to test if individuals satisfy the main implications of mixed strategy equilibrium in soccer penalty kicks. An important advantage of my dataset is that it contains multiple observations on specific heterogeneous pairs of players, a situation that rarely repeats in real life. I also study the effects of ignoring heterogeneity in empirical work, which arises when pooling observations because of the lack of repeated observations for specific pairs. I find that if researchers ignore heterogeneity when it is present, they may often reject the null when in fact the null is true. For that reason, I suggest a simple way of combining the test statistic of a player across opponents to obtain a valid aggregate test without making any additional assumptions.

From the empirical point of view, I find that the behavior of most soccer players, when repeatedly facing the same opponents, is consistent with the implications of mixed strategy equilibrium, in the sense that winning probabilities are identical across strategies, except for the kickers from the least professional team, and that player's actions are serially independent. In contrast, I find dependence between the kicker's and goalkeeper's actions. Moreover, the goalkeepers of the least professional team tended to replicate each other's action during the penalty shootout. Nevertheless, I also find that players do not seem to follow a reinforcement learning model.

Although the empirical analysis of this chapter provides reliable evidence on some fundamental implications of game theory, paying particular attention to the effects of the different years of experience and the level of professionalism of the different teams, there is still much to learn about the competitive behavior that arises in zero-sum games from the field, lab and real life situations.

2.8 Proofs and Auxiliary Results

2.8.1 *Proof of Proposition 2.1*

Recall from section 2.3 that under the null, $\pi^i_{hj} = \pi^i_h \times \pi^i_j$ for $h = S, F$ and $j = L, R$, where S, F, L and R are dummy variables with $F = 1 - S$ and $R = 1 - L$. The payoffs of a heterogeneous player H will be:

$$\begin{aligned}\pi^H_{hj} = P(h^H = 1, j^H = 1) = P(h^H = 1, j^H = 1 \mid Pair\ A) \times P(Pair\ A) \\ + P(h^H = 1, j^H = 1 \mid Pair\ B) \times P(Pair\ B)\end{aligned}$$

or equivalently

$$\pi^H_{hj} = \pi^A_{hj}\lambda + \pi^B_{hj}(1-\lambda).$$

Similarly,

$$\pi^H_m = P(m^H = 1 \mid Pair\ A) \times P(Pair\ A) + P(m^H = 1 \mid Pair\ B) \times P(Pair\ B)$$

or equivalently

$$\pi^H_m = \pi^A_m\lambda + \pi^B_m(1-\lambda), \text{ for } m = j, h.$$

We want to check if $\pi^i_{hj} = \pi^i_h \times \pi^i_j$ is true for the heterogeneous player H given that $\pi^A_{hj} = \pi^A_h \times \pi^A_j$ and $\pi^B_{hj} = \pi^B_h \times \pi^B_j$.

If we regress h^H on a constant and j, for example, the regression coefficient is

$$Cov(h^H, j^H)/Var(j^H),$$

where

$$Cov(h^H, j^H) = E(h^H j^H) - E(h^H)E(j^H) \text{ and } Var(j^H) = E\left[(j^H)^2\right] - \left[E(j^H)\right]^2.$$

Here,

$$E(h^H j^H) = E(h^A j^A \mid PairA)\lambda + E(h^B j^B \mid PairB)(1-\lambda),$$

but under independence of (h^A, j^A) and (h^B, j^B), then $E(h^H j^H) = (\pi^A_h \pi^A_j)\lambda + (\pi^B_h \pi^B_j)(1-\lambda)$ because $E(h^i j^i \mid Pair\ i) = E(h^i \mid Pair\ i) \times E(j^i \mid Pair\ i)$.

Similarly,

$$E(m^H) = \pi^A_m\lambda + \pi^B_m(1-\lambda), \text{ for } m = j, h.$$

Therefore,

$$Cov(h^H, j^H) = (\pi_h^A \pi_j^A)\lambda + (\pi_h^B \pi_j^B)(1-\lambda) - [\pi_j^A \lambda + \pi_j^B(1-\lambda)] \left[\pi_h^A \lambda + \pi_h^B(1-\lambda)\right],$$

which simplifies to

$$Cov(h^H, j^H) = \lambda(1-\lambda) \times \left(\pi_h^A - \pi_h^B\right) \times (\pi_j^A - \pi_j^B).$$

As a consequence, the regression coefficient will be zero if and only if $\lambda = 0$ or $\lambda = 1$ (no heterogeneity in the sample), or $\pi_h^A = \pi_h^B$ (no heterogeneity in the outcome) or $\pi_j^A = \pi_j^B$ (no heterogeneity in the strategies), as stated.

2.8.2 *Proof of lack of pure strategies*

Note that the penalty kick game in section 2.2.2 is strictly competitive because the kicker wants to score while the goalkeeper has opposite preferences (see Osborne (2003) chapter 11, section 3 for more details).

Let's find out the best response function of the goalkeeper to the actions of the kicker. If the kicker plays left, the goalkeeper best response is to play left because $1-a > 1-b$. Similarly, if the kicker plays center, the goalkeeper best response is to play center as $1-c > 1-e > 1-b$. Lastly, if the kicker plays right, the goalkeeper best response is to also play right. Therefore, the goalkeeper's best response is to play the same action as the kicker.

Now, let's derive the best response function of the kicker. If the goalkeeper plays left, the kicker best response is to play either center or right because $e > a$. Similarly, if the goalkeeper plays center, the kicker best response is to play left as $b > e > c$. Lastly, if the goalkeeper plays right, the kicker best response is to play left. Therefore, the kicker's best response is to play the opposite action of the goalkeeper.

2.8.3 *Existence and uniqueness of equilibrium*

Suppose that in the penalty kick game in section 2.2.2 the kicker believes that the goalkeeper plays L with probability q_L, R with probability q_R and C with probability $q_C = 1 - q_L - q_R$. Similarly, suppose the goalkeeper believes that the kicker plays L with probability p_L, R with probability p_R and C with probability $p_C = 1 - p_L - p_R$.

The expected payoff of the kicker's pure strategies against the goalkeeper's mixed strategies

(σ_G) for the payoff matrix in section 2.2.2 are:

$$\left.\begin{array}{l} E[u_K(L,\sigma_G)] = aq_L + b(q_C+q_R) \\ E[u_K(C,\sigma_G)] = e(q_L+q_R) + cq_C \\ E[u_K(R,\sigma_G)] = e(q_L+q_C) + dq_R \end{array}\right\}.$$

Since in equilibrium $E[u_K(L,\sigma_G)] = E[u_K(C,\sigma_G)] = E[u_K(R,\sigma_G)]$, then

$$\begin{aligned} aq_L + b(q_C+q_R) &= e(q_L+q_R) + cq_C \\ aq_L + b(q_C+q_R) &= e(q_L+q_C) + dq_R \end{aligned}$$

or equivalently

$$\begin{aligned} &(b+e-a-c)\,q_L + (e-c)\,q_R = b-c \\ &(a-b)\,q_L + (e-d)\,q_R = e-b \end{aligned}$$

which can be written in matrix form as

$$\begin{pmatrix} b+e-a-c & e-c \\ a-b & e-d \end{pmatrix}\begin{pmatrix} q_L \\ q_R \end{pmatrix} = \begin{pmatrix} b-c \\ e-b \end{pmatrix}.$$

Solving for q_L and q_R yields:

$$\begin{aligned} \begin{pmatrix} q_L \\ q_R \end{pmatrix} &= \begin{pmatrix} b+e-a-c & e-c \\ a-b & e-d \end{pmatrix}^{-1}\begin{pmatrix} b-c \\ e-b \end{pmatrix} \\ \begin{pmatrix} q_L \\ q_R \end{pmatrix} &= \frac{1}{\Delta}\begin{pmatrix} e-d & c-e \\ b-a & b+e-a-c \end{pmatrix}\begin{pmatrix} b-c \\ e-b \end{pmatrix} \end{aligned}$$

where $\Delta = (b+e-a-c)\,(e-d) - (e-c)\,(a-b) > 0$ given that $b > e > a > d > c$.

Hence,

$$\begin{pmatrix} q_L \\ q_R \end{pmatrix} = \frac{1}{\Delta}\begin{bmatrix} (e-d)\,(b-c) + (c-e)\,(e-b) \\ (e-a)\,(e-c) \end{bmatrix}.$$

Similarly, the expected payoff of the goalkeeper's pure strategies against the kicker's mixed strategies (σ_K) are:

$$\left.\begin{array}{l} E[u_G(L,\sigma_K)] = (1-a)p_L + (1-e)(p_C+p_R) \\ E[u_G(C,\sigma_K)] = (1-b)p_L + (1-e)p_R + (1-c)p_C \\ E[u_G(R,\sigma_K)] = (1-b)p_L + (1-e)p_C + (1-d)\,p_R \end{array}\right\}.$$

Since in equilibrium, $E[u_G(L,\sigma_K)] = E[u_G(C,\sigma_K)] = E[u_G(R,\sigma_K)]$, then

$$(1-a)p_L + (1-e)(p_C + p_R) = (1-b)p_L + (1-e)p_R + (1-c)p_C$$
$$(1-a)p_L + (1-e)(p_C + p_R) = (1-b)p_L + (1-e)p_C + (1-d)\,p_R$$

or equivalently

$$(e+b-a-c)\,p_L + (e-c)\,p_R = e-c$$
$$(a-b)\,p_L + (e-d)\,p_R = 0$$

which can be written in matrix form as

$$\begin{pmatrix} e+b-a-c & e-c \\ a-b & e-d \end{pmatrix}\begin{pmatrix} p_L \\ p_R \end{pmatrix} = \begin{pmatrix} e-c \\ 0 \end{pmatrix}$$

Solving for p_L and p_R yields:

$$\begin{pmatrix} p_L \\ p_R \end{pmatrix} = \frac{1}{\Delta}\begin{pmatrix} e-d & c-e \\ b-a & e+b-a-c \end{pmatrix}\begin{pmatrix} e-c \\ 0 \end{pmatrix}.$$

Hence,

$$\begin{pmatrix} p_L \\ p_R \end{pmatrix} = \frac{1}{\Delta}\begin{pmatrix} (e-d)\,(e-c) \\ (b-a)\,(e-c) \end{pmatrix}.$$

When $b > e > a > c$ but $c = d$, then $q_L > q_C = q_R$ and $p_L > p_C = p_R$ if $2b - d > 2e - a$, which implies that both players will choose more frequently the natural side of the kicker than any other side. In turn, the kicker will only choose R and the goalkeeper L when $b > e$ but $e = a = d$ because $p_L = p_C = q_C = q_R = 0$. Similarly, when $b > e$ and $a > d > c$ but $e = a$, the goalkeeper will only choose L because $q_C = q_R = 0$. Additionally, when $b = e = a$ but $a > d$, both players will only play L because $p_C = p_R = q_C = q_R = 0$. Finally, when $a = b = c = d = e$, both the kicker and the goalkeeper are indifferent between playing pure or mixed strategies, since the expected payoff from choosing L, C or R gives the exact same payoff. However, all these equilibria are ruled out by assumption.

Hence, the game has no pure strategy Nash equilibrium, as stated.

2.8.4 *Finite sample behavior under heterogeneity*

The p-value plots of the tests, which contain the empirical cumulative distribution function (cdf) of the asymptotic p-values in the Monte Carlo simulations for 200 observations (see Davidson and MacKinnon (1998)), are depicted in Figures 2.3.a to 2.3.f. The advantage of this sample size is that the asymptotic p-values are reliable. As can be seen, the empirical cdf of the asymptotic p-values is well above the 45° line even though both pairs of players play according to the theory. Hence, under heterogeneity, we will often mistakenly reject the null even though the null is true, as shown in Figure 2.2.

I also consider the case for 20 observations, where I get analogous results but with lower rejection rates. Thus, my Monte Carlo results confirm the result in Proposition 2.1.

2.8.5 *Size experiments*

Even in experimental studies, few observations for each pair of kicker-goalkeeper are likely to be the rule rather than the exception. Therefore, it is important to investigate the behavior of the tests described in section 2.5 in small samples because the asymptotic $\mathcal{X}^2$ distribution of those test procedures may be unreliable when the number of observations is small. Part of the problem is that given that all the variables used are discrete, the number of states of the world is finite (3 possible actions per player $\times$ 2 possible outcomes per combination of kicker and goalkeeper actions). In addition, the number of values of the estimators and test statistics will be repeated in many of those states of the world.

As we will see in chapter 3, there are only seven possible tests: the LM in the multivariate regression, which coincide with Pearson's independence test and the LM test in a multinomial model, multinomial logit and multinomial probit models as well as the J-test for overidentifying restrictions; the LR and Wald tests in the multivariate regression, Wald's heteroskedasticity-robust version, which coincides with the Wald test in the multinomial model, and the Wald and LR tests in the multinomial logit model, the last one being equal to the LR test in the probit and multinomial model. I will also consider the F-test of the univariate regression as the penalty kick has only two outcomes.

2.8.5.1 Problematic cases

Sometimes the calculations for some of the tests mentioned above breakdown. Although this is unlikely to happen with real data, I discuss here those situations in the 2×2 as an illustration because they occur in the simulations.

Perfect classification

As an example suppose that when the kicker shoots to the left, he never scores, whereas when he shoots in any other direction he may score or not. Hence, $\hat{\delta}_L = 0$ but $0 < \hat{\delta}_R < 1$. In this case, $\hat{\gamma}_L \rightarrow -\infty$ and the computation of the logit model breaks down (see Ruud (2000) section

27.1 for more details). The logit LR test is well defined although the unrestricted likelihood may also lead to numerical errors. For instance, when $\hat{\delta}_L \to 0$ the limit of the logit log-likelihood function becomes:

$$\lim_{\hat{\delta}_L \to 0} \log \left[\hat{\delta}_L^{\frac{n_{SL}}{n}} (1-\hat{\delta}_L)^{\frac{n_{FL}}{n}} \hat{\delta}_R^{\frac{n_{SR}}{n}} (1-\hat{\delta}_R)^{\frac{n_{FR}}{n}} \right]^n = \log \left[\hat{\delta}_R^{n_{SR}} (1-\hat{\delta}_R)^{n_{FR}} \right].$$

In this context, Stata removes the perfectly classified observations and computes again the MLE from the remaining ones. However, it fails to provide a Wald test. In that regard, I can prove that the limit of the logit Wald test goes to zero when one of the $\hat{\gamma}_i$, for $i = L, R$, goes to plus or minus infinity.

Perfect fit

In this case, the variables L and R explain perfectly the model, i.e. $R^2 = 1$. This requires $\hat{\delta}_L = 1$ and $\hat{\delta}_R = 0$ or vice versa. In this context, I can show that the LM test in the LPM is exactly equal to the number of observations, while the usual Wald, F and LR test as well as the heteroskedasticity-robust version of the Wald test of this regression model diverge to infinity. In the logit model, the LR can still be computed and it is not generally infinity, but the limit of the Wald test is surprisingly equal to 0.

The fact that the logit Wald test is 0 while the robust and non-robust versions of the Wald test in the LPM diverge to infinity confirms that this type of test is not numerically invariant to non-linear transformation of the restrictions (see again Ruud (2000) section 17.4 for another example in which two Wald tests based on transformation of the restrictions diverge).

Single outcome

This case arises when the estimated probability of scoring ($\hat{\pi}_S$) is either 0 or 1. This implies that the residual sum of squares of both the restricted and unrestricted model (SSR_R and SSR_U) are 0, which in turn implies that $\hat{\delta}_L = \hat{\delta}_R = 0$ or $\hat{\delta}_L = \hat{\delta}_R = 1$ depending on the value of $\hat{\pi}_S$.

When this occurs, I set all the tests for the LPM and logit models to 0, so that their p-values are 1.

Single choice

This occurs when the estimated probability of choosing left ($\hat{\pi}_L$) is either zero or one, which means that the player is only employing one strategy. When this case arises, I again set all the tests to 0 because the single choice situation is like the single outcome situation in the regression of L on a constant and $\tilde{y}$. Although, the theoretical results in section 2.3 show that $\pi_L = 0$ would not be optimal, $\hat{\pi}_L = 0$ can happen despite $\pi_L > 0$ if n is small.

Finally, there exists also the possibility that both the Single Choice and Single Outcome cases occur simultaneously, in which case I again set all the tests to 0.

2.8.5.2 Comparison asymptotic and Monte Carlo size in 2×3 and 3×3 Cases

I compare Monte Carlo and asymptotic p-values using p-value plots (see Davidson and MacKinnon (1998)), which display the empirical cdf of the asymptotic p-values in the Monte Carlo simulations.

I have simulated 10,000 replications of the 3×3 model explained in section 2.2.2 with $n = 20$ and parameter values $a = 0.03$, $b = 0.99$, $c = 0.01\dot{9}$, $d = 0.02$ and $e = 0.98$ to check if the p-value plots are close to the 45° degree line. Given that there are many values for the tests, I focus on p-values below 15%, which are the most relevant ones. The confidence intervals for the rejection rates at the 1, 5 and 10% levels under the null are $(0.80, 1.20)$, $(4.57, 5.43)$ and $(9.41, 10.6)$, respectively. In fact, the formula for calculating the confidence interval is $\alpha \pm 1.96\sqrt{\alpha(1-\alpha)/(\text{No. Replications})}$, where α is the significance level.

Scoring equality

The graphs for the null hypotheses of equal scoring probability are presented in Figures 2.4.a to 2.4.g. Empirical researchers that rely on asymptotic p-values should probably use the LR, LM and F-test of the LPM and avoid the rest of the tests. In particular, the Wald test in the LPM over rejects the null considerably.

Serial independence

The p-value plots corresponding to the null hypothesis of serial independence are presented in Figures 2.5.a to 2.5.h. Researchers that rely on asymptotic p-values should probably use F-versions of the Lawley-Hotelling (LH) and Wilks' test in the multivariate LPM, although the former is slightly better. At the same time, they should avoid the remaining tests. Again, the Wald test in the LPM and the multinomial logit LR test show considerable over rejections, while the Wald test in the multinomial logit hardly ever rejects.

Action independence

The graphs for the null hypotheses of action independence are presented in Figures 2.6.a to 2.6.h. The results suggest that empirical researches should probably use the LM test as well as the F-version of the Wilks, Lawley-Hotelling and Pillai tests in the multivariate LPM, with the

Wilks' test being almost perfect, and avoid the rest. In particular, the Wald and LR tests in the LPM and the multinomial logit LR test over reject the null considerably, while the Wald test in this model hardly rejects.

2.8.6 *Power experiments*

When choosing the significance level of a test, one sets the probability of rejecting the null when in fact it is true (Type 1 error). In the previous subsection, I studied which tests are more reliable when one chooses this level to be small, say 5%. At the same time, one would like to reject the null with high probability when the null is false. This is known as the power of the test. In what follows, I use Monte Carlo simulations to investigate the power of the different tests in some reasonable designs which do not satisfy the null.

2.8.6.1 Alternatives to equal scoring probabilities in 2×3 case

The alternative of the implication of equal scoring probabilities is that the player's probability of scoring depends on the strategy chosen. Here, I consider two alternatives because in reality, players do not necessarily know how to solve for the mixed strategy Nash equilibrium. At the same time, I have assumed serial independence to concentrate only in these two alternatives.

Alternative 1

In the case of the model with three actions, I assume that the kicker plays left with probability 0.7 and center and right with 0.15, while the goalkeeper plays left and center with probability 0.15 and right with 0.7. I have chosen these probabilities because, under the assumption that the kicker is left-footed, his natural side is to shoot to the left-hand side of the goalkeeper. Therefore, one could think that a naive kicker is more likely to shoot to the left, which justifies the kicker's probability of 0.7. However, the goalkeeper may believe that the obvious reasoning of a right-footed kicker is that the goalkeeper will jump to the left and therefore the probability of scoring will be low, so he will change the direction and kick to the right. That is why the chosen goalie's probability is 0.15.

Alternative 2

This alternative is similar to the previous one but now I assume that the kicker plays left and center with probability 0.15 and right with 0.7 while the goalkeeper plays left with probability 0.7 and center and right with 0.15.

Power of tests

I have only looked at the power of the LM and F tests of the null hypothesis of equal scoring probabilities in the LPM because they are the only ones whose asymptotic p-values are reliable under the null.

The following table shows the percentage of times that these tests reject the null at the 1, 5 and 10% significance level under alternatives 1 and 2.

	F-test		LM test	
%	Alternative 1	Alternative 2	Alternative 1	Alternative 2
1	22.42	24	18.25	19.64
5	36.56	38.63	36.28	38.2
10	44.89	47.27	45.47	47.86

These results suggest that empirical researchers should use the F-test of the LPM because it is the most powerful test under the two alternatives. The power here is higher than with only two actions because the alternatives are further away from the null.

2.8.6.2 Alternatives to serial independence in 3×3 contingency tables

The alternative of the hypothesis of serial independence is that the players actions at time t depend on the action chosen at time $t-1$. To propose a specific alternative, I use a standard Markov Chain (see section 3.4.1 for more details, and Shachat et al (2015) for a more complicated hidden Markov model). The transition matrix P of the Markov chain in this case is:

$$P = \begin{pmatrix} P_{LL} & P_{LC} & P_{LR} \\ P_{CL} & P_{CC} & P_{CR} \\ P_{RL} & P_{RC} & P_{RR} \end{pmatrix},$$

with states $i = L, C, R$, where $P_{iC} = 1 - P_{iL} - P_{iR}$.

Note that I can write the multivariate LPM to detect serial dependence as the following vector autoregression:

$$\left.\begin{aligned} L_t - \pi_L &= b_{LL}(L_{t-1} - \pi_L) + b_{RL}(R_{t-1} - \pi_R) + u_{Lt} \\ R_t - \pi_R &= b_{LR}(L_{t-1} - \pi_L) + b_{RR}(R_{t-1} - \pi_R) + u_{Rt} \end{aligned}\right\},$$

where $E(L_t) = \pi_L$ and $E(R_t) = \pi_R$ are the average probabilities of kicking left and right respectively. Assume for simplicity that $b_{RL} = b_{LR} = 0$ under the alternative. Hence I get that $P_{RL} = P_{CL} = P_{LL} - b_{LL}$ and $P_{LR} = P_{CR} = P_{RR} - b_{RR}$. These assumptions imply that the transition matrix is:

$$P = \begin{pmatrix} P_{LL} & 1 - P_{LL} - P_{RR} + b_{RR} & P_{RR} - b_{RR} \\ P_{LL} - b_{LL} & 1 - P_{LL} - P_{RR} + b_{LL} + b_{RR} & P_{RR} - b_{RR} \\ P_{LL} - b_{LL} & 1 - P_{LL} - P_{RR} + b_{LL} & P_{RR} \end{pmatrix}.$$

The stationary distribution in this Markov Chain is defined by the vector $\pi = (\pi_L, \pi_C, \pi_R)$, where $\pi_C = 1 - \pi_L - \pi_R$, such that $\pi P = \pi$.

This yields:

$$\pi_L = \frac{P_{LL} - b_{LL}}{1 - b_{LL}} \text{ and } \pi_R = \frac{P_{RR} - b_{RR}}{1 - b_{RR}}.$$

Under the null hypotheses, b_{LL} and b_{RR} are equal to zero. Here, I propose different alternatives depending on the values of b_{LL} and b_{RR}. However, I have assumed simultaneous moves and fixed the stationary probabilities of the Markov chain equal to the probabilities of the optimal strategy in the Nash equilibrium, so that the only discrepancy from the null is serial dependence.

Power of tests

I have only looked at the power of the F-version of the Lawley-Hotelling and Wilks' tests of the null hypothesis of serial independence in the multivariate LPM because they are the only ones whose asymptotic p-values are reliable under the null.

The following tables show the percentage of times that these tests reject the null at the 1, 5 and 10% significance level under the following alternatives: (1) $b_{LL} = b_{RR} = 0.5$, (2) $b_{LL} = b_{RR} = -0.5$, (3) $b_{LL} = b_{RR} = 0.05$ and (4) $b_{LL} = b_{RR} = -0.05$.

	Wilks' F-test				Lawley-Hotelling F-test			
%	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
1	23.30	50.67	1.08	1.49	24.55	50.82	1.37	1.84
5	40.37	92.31	4.42	5.64	40.92	86.03	4.79	6.03
10	50.65	97.48	8.48	10.60	51.25	97.09	8.74	10.92

These results suggest that empirical researchers should use the F-version of the Lawley-Hotelling test in the multivariate LPM because it is the most powerful test under the four alternatives, although it does not have much power under alternatives 3 and 4, as expected.

2.8.6.3 Alternatives to action independence in 3×3 case

Taking into account the discussion in section 2.5.3, I can test the null hypothesis by estimating the following model:

$$\left.\begin{array}{l} L_{Gt} = \xi_{L0} + \xi_{LL} L_{Kt} + \xi_{LR} R_{Kt} + u_{Lt} \\ R_{Gt} = \xi_{R0} + \xi_{RL} L_{Kt} + \xi_{RR} R_{Kt} + u_{Rt} \end{array}\right\},$$

where L_{Gt} and R_{Gt} is the direction chosen by the goalkeeper at time t and L_{Kt} and R_{Kt} is the direction chosen by the kicker at time t.

Note that this model can be written as:

$$\left.\begin{array}{l} L_{Gt} - \pi_{GL} = \xi_{LL}(L_{Kt} - \pi_{KL}) + \xi_{LR}(R_{Kt} - \pi_{KR}) + u_{Lt} \\ R_{Gt} - \pi_{GR} = \xi_{RL}(L_{Kt} - \pi_{KL}) + \xi_{RR}(R_{Kt} - \pi_{KR}) + u_{Rt} \end{array}\right\},$$

where π_{hj}, for $h = K, G$ and $j = L, R$, denote the average probabilities of a player's strategy and ξ_{ij}, for $i, j = L, R$ denotes a correlation measure between the actions of the kicker and goalkeeper. If the ξ coefficients are not 0, then either the goalie anticipates the action of the kicker or the kicker misleads the goalkeeper. I set the values of $\pi_{Kj} = p_j$ and $\pi_{Gj} = q_j$ so that on average the players satisfy the implications of mixed strategy equilibrium but not at any particular penalty kick.

Under the null hypotheses, ξ_{LL}, ξ_{LR}, ξ_{RL} and ξ_{RR} are equal to 0. For simplicity, I assume that $\xi_{LR} = \xi_{RL} = 0$ and consider different alternatives depending on the values of ξ_{LL} and ξ_{RR}.

Power of tests

I have only looked at the power of the LM test as well as the F-versions of the Wilks, LH, Pillai tests in the multivariate LPM because they are the only ones which performed well under the null.

The following table shows the percentage of times that the above tests reject the null at the 1, 5 and 10% significance level under the following alternatives: (1) $\xi_{LL} = \xi_{RR} = 0.5$, (2) $\xi_{LL} = \xi_{RR} = -0.5$, (3) $\xi_{LL} = \xi_{RR} = 0.05$ and (4) $\xi_{LL} = \xi_{RR} = -0.05$.

	LM test				Lawley-Hotelling F-test			
%	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
1	40.45	28.34	0.67	0.66	48.10	43.58	2.05	1.87
5	66.10	95.14	5.20	4.64	66.62	86.75	6.50	6.41
10	76.43	98.10	11.04	11.02	75.39	97.51	11.08	11.16

	Pillai F-test				Wilks F-test			
%	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
1	44.56	36.32	0.99	0.88	47.16	42.24	1.61	1.38
5	66.01	94.98	5.15	4.59	66.37	92.19	5.99	5.58
10	75.43	97.99	10.27	10.14	75.60	97.67	10.84	10.79

These results suggest that empirical researchers should use the F version of the Lawley-Hotelling trace test in the multivariate LPM because it is the most powerful test under the four alternatives, although as expected, it does not have much power under alternatives 3 and 4.

2.8.7 *Multiple testing issues*

With N players in my sample, there will be N aggregate test statistics, and some of those tests could reject the null by chance even when the null is true for all of them. In fact, the probability of k of N tests rejecting the null is given by the following binomial distribution:

$$\Pr(X = k) = \binom{N}{k} p^k (1-p)^{N-k}. \text{ for } k = 0, \ldots, N, \tag{2.6}$$

where $\binom{N}{k} = N!/[k!\,(N-k)]$, N is the total number of players considered, k is the number of rejections and p is the significance level of the tests.

In empirical work, researchers often compare the actual number of rejections to the expected number of rejections, but this rule is only reliable when the number of players N is very large. For that reason, I will use the binomial probabilities (2.6) to confirm whether players treated as a group behave consistently with the theory.

2.9 Tables Chapter 2

Table 2.1: Distribution of Strategies Combinations and Scoring Rates

	#Obs.	*L-L*	*L-C*	*L-R*	*C-L*	*C-C*	*C-R*	*R-L*	*R-C*	*R-R*	Scoring Rate
Cadete A	129	27.13	3.87	14.73	3.87	4.65	4.65	21.70	3.13	16.27	72.86
Cadete C	108	31.49	1.85	17.6	2.77	2.77	1.85	22.22	3.70	15.75	77.77
Juvenil A	104	16.34	0	25	8.65	1.92	3.84	13.46	3.84	26.95	76.92
All penalties	341	25.22	2.05	18.76	4.98	3.22	3.51	19.35	3.51	19.40	75.66
Scoring Rate	258	60.46	71.42	92.18	82.35	36.36	91.66	95.45	100	57.57	

Notes: The first letter refers to the choice made by the kicker (Left (L), Center (C) and Right (R)) and the second one to the choice made by the goalkeeper, always from the point of view of the goalkeeper. For instance, L-R means than the kicker chooses to kick to the left hand side of the goalkeeper and the goalkeeper chooses to jump to his right.

Table 2.2: Test of Equal Scoring Probabilities Cadete A

Pair #	Player	#Obs.	Frequency			Scoring Rate			F-Test	p-value
			L	C	R	L	C	R		
1	Kicker	16	50	25	25	0.75	0.25	0.75	1.63	0.24
	Goalkeeper		37.50	25	37.50	0.66	0.25	0.83	1.86	0.20
2	Kicker	13	69.23	7.69	23.08	0.77	0	1	2.42	0.14
	Goalkeeper		53.85	23.08	23.08	0.71	0.66	1	0.51	0.62
3	Kicker	9	33.33	11.11	55.56	0.66	1	0.80	0.18	0.84
	Goalkeeper		77.78	-	22.22	0.85	-	0.50	1.02	0.35
4	Kicker**	25	20	32	48	0.20	0.87	0.83	5.59	0.01
	Goalkeeper		56	16	28	0.78	0.83	0.66	0.58	0.57
5	Kicker	25	56	4	40	0.57	1	0.70	0.46	0.64
	Goalkeeper		48	8	44	0.58	0.50	0.72	0.32	0.73
6	Kicker	8	75	-	25	1	-	1	0	1
	Goalkeeper		87.50	-	12.50	1	-	1	0	1
7	Kicker	14	50	-	50	0.71	-	0.71	0	1
	Goalkeeper**		42.86	7.14	50	0.33	1	1	6.29	0.02
8	Kicker	19	36.84	10.53	52.63	1	1	0.60	2.53	0.11
	Goalkeeper*		47.37	5.26	47.37	1	0.55	1	3.37	0.06

Notes: L (Left), C (Center) and R (Right) denote the strategies available to the players. Additionally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

Table 2.3: Test of Equal Scoring Probabilities Cadete C

Pair #	Player	#Obs.	Frequency			Scoring Rate			F-Test	p-value
			L	C	R	L	C	R		
1	Kicker	13	61.54	15.38	23.08	0.75	0.25	0.75	0.64	0.55
	Goalkeeper		30.77	15.38	53.85	0.75	0.25	0.83	0.26	0.77
2	Kicker	14	57.14	7.14	35.71	0.75	0	1	0.14	0.87
	Goalkeeper		50	14.29	35.71	0.71	0.66	1	0.32	0.74
3	Kicker**	13	30.77	-	69.23	0.50	1	0.80	7.62	0.02
	Goalkeeper		92.31	-	7.69	0.83	-	0.50	0.17	0.69
4	Kicker	14	42.86	-	57.14	0.83	-	0.75	0.12	0.73
	Goalkeeper*		50	-	50	1	-	0.57	4.50	0.05
5	Kicker**	14	57.14	7.14	35.71	0.87	0	1	5.28	0.02
	Goalkeeper***		21.43	7.14	71.43	0.66	-	1	8.64	0
6	Kicker	8	69.23	7.69	23.08	0.66	1	0.66	0.19	0.83
	Goalkeeper		61.54	7.69	30.77	0.75	1	0.5	0.54	0.60
7	Kicker	14	35.71	21.43	42.86	0.60	0.33	0.83	1.05	0.38
	Goalkeeper		78.57	7.14	14.29	0.63	-	0.50	0.30	0.74
8	Kicker	13	53.85	-	46.15	0.71	-	0.83	0.22	0.65
	Goalkeeper		69.23	7.69	23.08	0.77	1	0.66	0.19	0.83

Notes: L (Left), C (Center) and R (Right) denote the strategies available to the players. Additionally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

Table 2.4: Test of Equal Scoring Probabilities Juvenil A

			Frequency			Scoring Rate				
Pair #	Player	#Obs.	L	C	R	L	C	R	F-Test	p-value
1	Kicker	13	23.08	7.69	69.23	1	1	0.77	0.44	0.66
	Goalkeeper		7.69	23.08	69.23	1	1	0.77	0.44	0.66
2	Kicker	20	25	20	55	0.80	0.75	0.75	0.04	0.96
	Goalkeeper		25	5	70	0.60	1	0.78	0.46	0.64
3	Kicker	6	33.33	-	66.67	0.50	-	1	2.67	0.18
	Goalkeeper		33.33	-	66.67	0.50	-	1	2.67	0.18
4	Kicker	13	46.15	7.69	46.15	1	1	1	0	1
	Goalkeeper		61.54	7.69	30.77	1	1	1	0	1
5	Kicker	20	70	10	20	0.71	0.50	0.25	1.43	0.27
	Goalkeeper		45	5	50	0.66	0	0.60	0.77	0.48
6	Kicker	6	50	-	50	1	-	0.33	4	0.12
	Goalkeeper		16.67	-	83.33	1	-	0.60	0.44	0.54
7	Kicker	6	66.67	-	33.33	0.75	-	1	0.44	0.54
	Goalkeeper		66.67	-	33.33	0.75	-	1	0.44	0.54
8	Kicker	7	28.57	14.29	57.14	1	1	1	0	1
	Goalkeeper		42.86	-	57.14	1	-	1	0	1
9	Kicker	6	50	33.33	16.67	0.66	1	1	0.37	0.72
	Goalkeeper		83.33	-	16.67	0.80	-	1	0.17	0.70
10	Kicker	7	14.29	57.14	28.57	0	0.75	0	2.57	0.19
	Goalkeeper		28.57	-	71.43	0.50	-	0.40	0.04	0.85

Notes: L (Left), C (Center) and R (Right) denote the strategies available to the players. Additionally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

Table 2.5: Test of Equal Scoring Probabilities with Heterogeneity

Panel A

Player Cadete A	Test Statistic	Degrees of Freedom	p-value
Kicker 1*	8.10	4	0.09
Kicker 2**	11.54	4	0.02
Kicker 3	0.92	3	0.82
Kicker 4	5.06	3	0.17
Goalkeeper 1**	17.96	7	0.01
Goalkeeper 2	8.92	7	0.26

Panel B

Player Cadete C	Test Statistic	Degrees of Freedom	p-value
Kicker 1	1.56	4	0.82
Kicker 2**	7.74	2	0.02
Kicker 3**	10.94	4	0.03
Kicker 4	2.32	3	0.51
Goalkeeper 1***	18.57	7	0
Goalkeeper 2	6.60	7	0.47

Panel C

Player Juvenil A	Test Statistic	Degrees of Freedom	p-value
Kicker 1	0.88	2	0.64
Kicker 2	2.75	3	0.43
Kicker 3	0	2	1
Kicker 4*	6.86	3	0.09
Kicker 5	0.44	3	0.93
Kicker 6	5.88	4	0.21
Goalkeeper 1	3.38	10	0.97
Goalkeeper 2	3.72	4	0.45

Notes: The test statistic for a given player is the sum of the independent $\mathcal{X}^2$ statistics across all his opponents with degrees of freedom equal to the sum of the degrees of freedom for each pair test. Additionally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

Table 2.6: Test for Serial Independence of Actions Cadete A

Pair #	Player	#Obs.	Transition Matrix									LH F-Test	p-value
			$L_t\|L_{t-1}$	$C_t\|L_{t-1}$	$R_t\|L_{t-1}$	$L_t\|C_{t-1}$	$C_t\|C_{t-1}$	$R_t\|C_{t-1}$	$L_t\|R_{t-1}$	$C_t\|R_{t-1}$	$R_t\|R_{t-1}$		
1	Kicker	14	0.42	0.29	0.29	1	0	0	0.25	0.25	0.25	0.99	0.44
	Goalkeeper		0.33	0.17	0.50	0.67	0	0.33	0.40	0.20	0.40	0.22	0.93
2	Kicker	11	0.57	0.14	0.29	0	0	1	1	0	0	0.95	0.47
	Goalkeeper		0.33	0.50	0.17	1	0	0	0.33	0	0.67	1.30	0.33
3	Kicker	7	0	0.50	0.50	1	0	0	0.25	0	0.75	0.81	0.58
	Goalkeeper		0.80	-	0.20	-	-	-	1	-	0	0.36	0.58
4	Kicker	23	0	0.60	0.40	0.29	0.29	0.42	0.18	0.27	0.55	0.60	0.67
	Goalkeeper		0.59	0.08	0.33	0.50	0.25	0.25	0.57	0.29	0.14	0.39	0.81
5	Kicker	23	0.64	0.07	0.29	0	0	1	0.37	0	0.63	0.94	0.45
	Goalkeeper		0.50	0.08	0.42	1	0	0	0.44	0.44	0.12	0.45	0.77
6	Kicker	6	0.80	-	0.20	-	-	-	0	-	1	2.67	0.18
	Goalkeeper		0.80	-	0.20	-	-	-	1	-	0	0.17	0.70
7	Kicker	12	0.33	-	0.67	-	-	-	0.50	-	0.50	0.29	0.60
	Goalkeeper		0.50	0	0.50	1	0	0	0.20	0.20	0.60	0.66	0.63
8	Kicker	17	0.17	0.17	0.66	0	0	1	0.60	0.10	0.30	0.93	0.47
	Goalkeeper		0.44	0.12	0.44	-	-	-	0.63	0	0.37	0.53	0.560

Notes: L_t (Left), C_t (Center) and R_t (Right) denote the strategies available to the players at time t while L_{t-1}, C_{t-1} and R_{t-1} are its corresponding lagged variables. For instance, $L_t|R_{t-1}$ means that the player chooses L at time t after choosing R at $t-1$. Additionally, LH F-test denotes the F-version of the Lawley-Hotelling test in the multivariate LPM. Finally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

Table 2.7: Test for Serial Independence of Actions Cadete C

Pair #	Player	#Obs.	Transition Matrix									LH F-Test	p-value
			$L_t\|L_{t-1}$	$C_t\|L_{t-1}$	$R_t\|L_{t-1}$	$L_t\|C_{t-1}$	$C_t\|C_{t-1}$	$R_t\|C_{t-1}$	$L_t\|R_{t-1}$	$C_t\|R_{t-1}$	$R_t\|R_{t-1}$		
1	Kicker	10	0.43	0.14	0.43	0	1	0	1	0	0	1.42	0.30
	Goalkeeper		0	0.25	0.75	0	0	1	0.60	0	0.40	1.19	0.37
2	Kicker	11	0.83	0	0.17	0	0	1	0.50	0.25	0.25	0.92	0.48
	Goalkeeper		0.60	0.20	0.20	1	0	0	0.25	0.25	0.50	0.63	0.65
3	Kicker	10	0.50	-	0.50	-	-	-	0.17	-	0.83	1.16	0.31
	Goalkeeper		0.11	0.89	-	0	1	-	-	-	-	0.10	0.76
4	Kicker	11	0.40	-	0.60	-	-	-	0.33	-	0.67	0.04	0.84
	Goalkeeper		0.40	-	0.60	-	-	-	0.67	-	0.33	0.69	0.43
5	Kicker	11	0.57	0	0.43	-	-	-	0.50	0.25	0.25	0.89	0.45
	Goalkeeper		0	0	1	-	-	-	0.11	0.11	0.78	0.21	0.82
6	Kicker	10	0.43	0.14	0.43	-	-	-	1	0	0	1.40	0.31
	Goalkeeper		0.43	0.14	0.43	-	-	-	0.67	0	0.33	0.27	0.77
7	Kicker	11	0.25	0.25	0.50	0.50	0	0.50	0.20	0.40	0.40	0.22	0.93
	Goalkeeper		0.75	0.13	0.13	1	0	0	0.50	0	0.50	0.33	0.85
8	Kicker	10	0.50	-	0.50	-	-	-	0.67	-	0.33	0.23	0.65
	Goalkeeper		0.72	0.14	0.14	-	-	-	0.67	0	0.33	0.31	0.74

Notes: L_t (Left), C_t (Center) and R_t (Right) denote the strategies available to the players at time t while L_{t-1}, C_{t-1} and R_{t-1} are its corresponding lagged variables. For instance, $L_t|R_{t-1}$ means that the player chooses L at time t after choosing R at $t-1$. Additionally, LH F-test denotes the F-version of the Lawley-Hotelling test in the multivariate LPM. Finally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

Table 2.8: Test for Serial Independence of Actions Juvenil A

Pair #	Player	#Obs.	Transition Matrix									LH F-Test	p-value
			$L_t\|L_{t-1}$	$C_t\|L_{t-1}$	$R_t\|L_{t-1}$	$L_t\|C_{t-1}$	$C_t\|C_{t-1}$	$R_t\|C_{t-1}$	$L_t\|R_{t-1}$	$C_t\|R_{t-1}$	$R_t\|R_{t-1}$		
1	Kicker	12	0	0.50	0.50	1	0	0	0.22	0	0.78	2.24	0.12
	Goalkeeper		0	0	1	0	0	1	0.12	0.38	0.50	1.50	0.27
2	Kicker	18	0.50	0	0.50	0.50	0	0.50	0.10	0.30	0.60	0.83	0.45
	Goalkeeper		0	-	1	0	-	1	0.33	-	0.67	1.25	0.32
3	Kicker	5	0.50	-	0.50	-	-	-	0.33	-	0.67	0.09	0.79
	Goalkeeper		0	-	1	-	-	-	0.50	-	0.50	0.60	0.49
4	Kicker**	12	0	0.17	0.83	1	0	0	0.80	0	0.20	4.81	0.01
	Goalkeeper		0.50	0.13	0.37	1	0	0	1	0	0	1.50	0.27
5	Kicker	18	0.67	0.08	0.25	1	0	0	0.75	0.25	0	0.55	0.70
	Goalkeeper		0.37	0	0.63	1	0	0	0.56	0.11	0.33	0.63	0.65
6	Kicker	5	1	-	0	-	-	-	0.33	-	0.67	2.40	0.22
	Goalkeeper		0	-	1	-	-	-	0.25	-	0.75	0.20	0.69
7	Kicker	6	0	0	1	-	-	-	0.50	0.25	0.25	0.60	0.49
	Goalkeeper		0	-	1	-	-	-	0.67	-	0.33	1.80	0.27
8	Kicker	5	0.50	-	0.50	-	-	-	1	-	0	1.50	0.35
	Goalkeeper		1	-	0	-	-	-	0.50	-	0.50	4	0.12
9	Kicker***	5	0.50	0	0.50	1	0	0	0	1	0	∞	0
	Goalkeeper		0.75	-	0.25	-	-	-	1	-	0	0.20	0.69
10	Kicker	6	-	1	0	-	0.33	0.67	-	1	0	1.33	0.31
	Goalkeeper		0	-	1	-	-	-	0.50	-	0.50	1.33	0.31

Notes: L_t (Left), C_t (Center) and R_t (Right) denote the strategies available to the players at time t while L_{t-1}, C_{t-1} and R_{t-1} are its corresponding lagged variables. For instance, $L_t|R_{t-1}$ means that the player chooses L at time t after choosing R at $t-1$. Additionally, LH F-test denotes the F-version of the Lawley-Hotelling test in the multivariate LPM. Finally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

Table 2.9: Test for Serial Independence of Actions with Heterogeneity

Panel A

Player Cadete A	Test Statistic	Degrees of Freedom	p-value
Kicker 1	8.74	8	0.36
Kicker 2	8.44	8	0.39
Kicker 3	6.40	5	0.27
Kicker 4	4.46	5	0.48
Goalkeeper 1	7.29	13	0.89
Goalkeeper 2	9.89	11	0.54

Panel B

Player Cadete C	Test Statistic	Degrees of Freedom	p-value
Kicker 1	11.63	8	0.17
Kicker 2	1.32	2	0.52
Kicker 3	4.85	4	0.30
Kicker 4	1.67	5	0.89
Goalkeeper 1	7.98	11	0.72
Goalkeeper 2	5.61	9	0.78

Panel C

Player Juvenil A	Test Statistic	Degrees of Freedom	p-value
Kicker 1*	8.69	4	0.07
Kicker 2	5.18	5	0.39
Kicker 3*	8.81	4	0.07
Kicker 4	4.99	5	0.42
Kicker 5	3.83	3	0.28
Kicker 6	9.67	6	0.14
Goalkeeper 1	16.16	16	0.44
Goalkeeper 2	3.33	4	0.50

Notes: The test statistic for a given player is the sum of the independent $\mathcal{X}^2$ statistics across all his opponents with degrees of freedom equal to the sum of the degrees of freedom for each pair test. Additionally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

Table 2.10: Test for Action Independence Cadete A

		Distribution of Strategies										
Pair #	#Obs.	*L-L*	*L-C*	*L-R*	*C-L*	*C-C*	*C-R*	*R-L*	*R-C*	*R-R*	LH F-Test	*p*-value
1*	16	0.25	0.06	0.19	0	0.19	0.06	0.37	0.25	0.37	2.58	0.06
2	13	0.38	0.08	0.23	0	0.08	0	0.15	0.08	0	1.16	0.36
3	9	0.33	-	0	0	-	0.11	0.44	-	0.11	2.83	0.14
4	25	0.16	0.04	0	0.16	0.08	0.08	0.24	0.04	0.2	0.85	0.50
5	25	0.32	0.04	0.2	0	0	0.04	0.16	0.04	0.2	0.43	0.78
6	8	0.62	-	0.12	-	-	-	0.25	-	0	0.30	0.60
7	14	0.21	0	0.29	-	-	-	0.21	0.07	0.21	0.49	0.63
8	19	0.16	0.05	0.16	0.05	0	0.05	0.26	0.05	0.26	0.37	0.83

Notes: The first letter refers to the choice made by the kicker (Left (L), Center (C) and Right (R)) and the second one to the choice made by the goalkeeper, always from the point of view of the goalkeeper. For instance, *R-L* means than the kicker chooses to kick to the right hand side of the goalkeeper and the goalkeeper chooses to jump to his left. Additionally, LH F-test denotes the F-version of the Lawley-Hotelling test in the multivariate LPM. Finally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

Table 2.11: Test for Action Independence Cadete C

		Distribution of Strategies										
Pair #	#Obs.	*L-L*	*L-C*	*L-R*	*C-L*	*C-C*	*C-R*	*R-L*	*R-C*	*R-R*	LH F-Test	*p*-value
1	13	0.15	0.08	0.38	0.08	0	0.08	0.08	0.08	0.08	0.28	0.89
2*	14	0.29	0	0.29	0	0.07	0	0.21	0.07	0.07	2.75	0.06
3	13	0.31	0	-	-	-	-	0.62	0.08	-	0.42	0.53
4	15	0.21	-	0.21	-	-	-	0.29	-	0.28	0	1
5***	15	0.21	0	0.36	0	0.07	0	0	0	0.36	∞	0
6***	14	0.54	0	0.15	0	0.08	0	0.08	0	0.15	∞	0
7	14	0.36	0	0	0.14	0	0.07	0.29	0.07	0.07	0.61	0.66
8*	13	0.46	0.08	0	-	-	-	0.23	0	0.23	3.08	0.09

Notes: The first letter refers to the choice made by the kicker (Left (L), Center (C) and Right (R)) and the second one to the choice made by the goalkeeper, always from the point of view of the goalkeeper. For instance, R-L means than the kicker chooses to kick to the right hand side of the goalkeeper and the goalkeeper chooses to jump to his left. Additionally, LH F-test denotes the F-version of the Lawley-Hotelling test in the multivariate LPM. Finally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

Table 2.12: Test for Action Independence Juvenil A

Pair #	#Obs.	Distribution of Strategies									LH F-Test	p-value
		L-L	*L-C*	*L-R*	*C-L*	*C-C*	*C-R*	*R-L*	*R-C*	*R-R*		
1	13	0	0	0.23	0	0.08	0	0.08	0.15	0.46	1.08	0.40
2**	20	0.1	0	0.15	0.15	0	0.05	0	0.05	0.5	3.63	0.02
3	6	0.17	-	0.17	0.08	-	-	0.17	-	0.5	0.27	0.63
4	13	0.23	0	0.23	0.05	0	0	0.31	0.08	0.08	0.55	0.70
5***	20	0.35	0	0.35	-	0.05	0	0.05	0	0.15	4.04	0
6	6	0	-	0.5	-	-	-	0.17	-	0.33	1	0.37
7	6	0.33	-	0.33	0	-	-	0.33	-	0	1.33	0.31
8	7	0	-	0.29	-	-	0.14	0.43	-	0.14	2.57	0.19
9	6	0.33	-	0.17	0.33	-	0	0.17	-	0	0.38	0.72
10	7	0	-	0.14	0.29	-	0.29	0	-	0.29	0.86	0.49

Notes: The first letter refers to the choice made by the kicker (Left (L), Center (C) and Right (R)) and the second one to the choice made by the goalkeeper, always from the point of view of the goalkeeper. For instance, *R-L* means than the kicker chooses to kick to the right hand side of the goalkeeper and the goalkeeper chooses to jump to his left. Additionally, LH F-test denotes the F-version of the Lawley-Hotelling test in the multivariate LPM. Finally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

Table 2.13: Test for Sequential Independence

Panel A

Player Cadete A	#Obs.	Transition Matrix									LH F-Test	p-value
		$L_t\|L_{t-1}$	$C_t\|L_{t-1}$	$R_t\|L_{t-1}$	$L_t\|C_{t-1}$	$C_t\|C_{t-1}$	$R_t\|C_{t-1}$	$L_t\|R_{t-1}$	$C_t\|R_{t-1}$	$R_t\|R_{t-1}$		
Kickers	127	0.46	0.17	0.37	0.25	0.12	0.63	0.52	0.08	0.40	1.42	0.23
Goalkeepers	127	0.55	0.11	0.34	0.53	0.20	0.27	0.49	0.09	0.42	0.58	0.67

Panel B

Player Cadete C	#Obs.	Transition Matrix									LH F-Test	p-value
		$L_t\|L_{t-1}$	$C_t\|L_{t-1}$	$R_t\|L_{t-1}$	$L_t\|C_{t-1}$	$C_t\|C_{t-1}$	$R_t\|C_{t-1}$	$L_t\|R_{t-1}$	$C_t\|R_{t-1}$	$R_t\|R_{t-1}$		
Kickers	105	0.45	0.10	0.45	0.50	0	0.50	0.57	0.07	0.36	0.51	0.73
Goalkeepers**	105	0.43	0.10	0.47	0.67	0	0.33	0.76	0.08	0.16	3.16	0.01

Panel C

Player Juvenil A	#Obs.	Transition Matrix									LH F-Test	p-value
		$L_t\|L_{t-1}$	$C_t\|L_{t-1}$	$R_t\|L_{t-1}$	$L_t\|C_{t-1}$	$C_t\|C_{t-1}$	$R_t\|C_{t-1}$	$L_t\|R_{t-1}$	$C_t\|R_{t-1}$	$R_t\|R_{t-1}$		
Kickers	102	0.34	0.15	0.51	0.47	0.13	0.40	0.48	0.13	0.39	0.45	0.77
Goalkeepers	102	0.39	0.05	0.56	0.17	0	0.83	0.40	0.07	0.53	0.54	0.70

Notes: L_t (Left), C_t (Center) and R_t (Right) denote the strategies available to the players at time t while L_{t-1}, C_{t-1} and R_{t-1} are its corresponding lagged variables. For instance, $L_t|R_{t-1}$ means that the player chooses L at time t after the previous player chose R at $t-1$. Additionally, LH F-test denotes the F-version of the Lawley-Hotelling test in the multivariate LPM. Finally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

Table 2.14: Reinforced Learning Cadete A

Pair #	Player	#Obs.	*LM* Test	*p*-value
1	Kicker	14	10.78	0.37
	Goalkeeper		11.67	0.31
2	Kicker	11	10.48	0.11
	Goalkeeper		10.39	0.24
3	Kicker*	7	14	0.08
	Goalkeeper		0.88	0.83
4	Kicker	23	8.53	0.58
	Goalkeeper		10.77	0.38
5	Kicker	23	6.88	0.55
	Goalkeeper		7.88	0.64
6	Kicker	6	2.40	0.12
	Goalkeeper		0.24	0.62
7	Kicker	12	0.69	0.88
	Goalkeeper		6.78	0.34
8	Kicker	17	7.03	0.32
	Goalkeeper		1.74	0.78

Notes: *Indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

Table 2.15: Reinforced Learning Cadete C

Pair #	Player	#Obs.	*LM* Test	*p*-value
1	Kicker	10	9.43	0.15
	Goalkeeper		7.22	0.51
2	Kicker	11	9.69	0.29
	Goalkeeper		8.56	0.38
3	Kicker**	10	6.03	0.04
	Goalkeeper		0.48	0.79
4	Kicker	11	3.44	0.33
	Goalkeeper		3.47	0.18
5	Kicker	11	2.98	0.56
	Goalkeeper		0.54	0.97
6	Kicker	10	8.17	0.23
	Goalkeeper*		11	0.09
7	Kicker	11	7.09	0.53
	Goalkeeper		8.25	0.41
8	Kicker	10	2.50	0.47
	Goalkeeper		9.29	0.16

Notes: *Indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

Table 2.16: Reinforced Learning Juvenil A

Pair #	Player	#Obs.	LM Test	p-value
1	Kicker	12	9.79	0.13
	Goalkeeper		6	0.42
2	Kicker	18	8.86	0.55
	Goalkeeper		5.14	0.27
3	Kicker	5	2.22	0.33
	Goalkeeper		0.83	0.36
4	Kicker*	12	8.81	0.07
	Goalkeeper		3	0.56
5	Kicker	18	6.49	0.77
	Goalkeeper		5.22	0.73
6	Kicker	5	2.92	0.23
	Goalkeeper		1.88	0.39
7	Kicker	5	2.22	0.33
	Goalkeeper		1.88	0.39
8	Kicker	6	3	0.22
	Goalkeeper*		3	0.08
9	Kicker	5	6.67	0.15
	Goalkeeper		0.31	0.58
10	Kicker	6	3.75	0.29
	Goalkeeper		3	0.39

Notes: *Indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

Table 2.17: Reinforced Learning under Heterogeneity

Panel A

Player Cadete A	Test Statistic	Degrees of Freedom	p-value
Kicker 1	21.26	16	0.17
Kicker 2	22.53	18	0.21
Kicker 3	9.28	9	0.41
Kicker 4	7.72	9	0.56
Goalkeeper 1	27.20	29	0.56
Goalkeeper 2	23.14	23	0.45

Panel B

Player Cadete C	Test Statistic	Degrees of Freedom	p-value
Kicker 1	19.12	14	0.16
Kicker 2*	9.47	5	0.09
Kicker 3	11.15	10	0.35
Kicker 4	9.59	11	0.57
Goalkeeper 1	16.49	22	0.79
Goalkeeper 2*	32.31	22	0.07

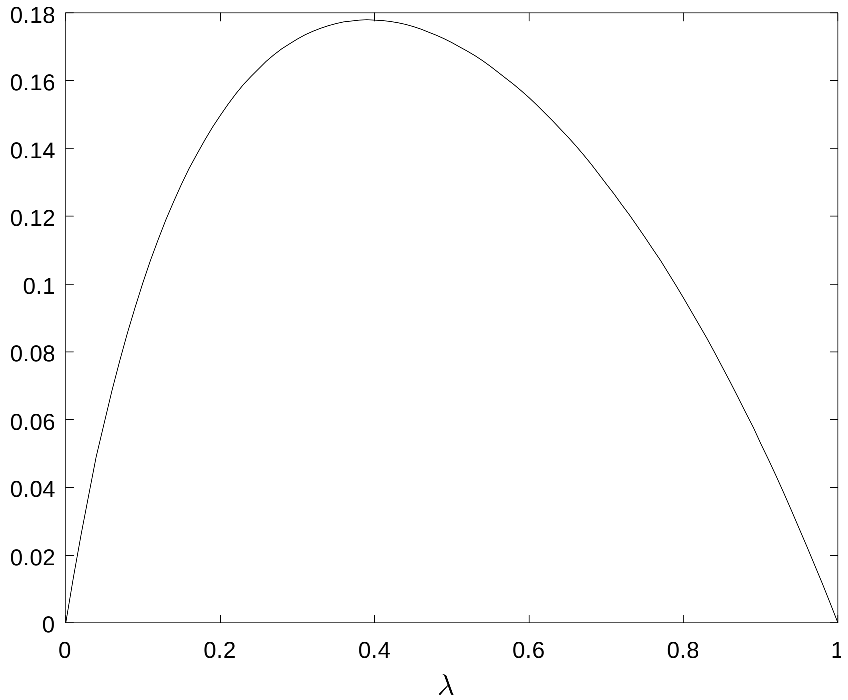
Panel C

Player Juvenil. A	Test Statistic	Degrees of Freedom	p-value
Kicker 1	9.79	6	0.13
Kicker 2	11.08	12	0.52
Kicker 3*	8.81	4	0.07
Kicker 4	9.41	12	0.67
Kicker 5	5.22	4	0.26
Kicker 6	10.42	7	0.17
Goalkeeper 1	25.36	26	0.50
Goalkeeper 2	4.89	6	0.56

Notes: The test statistic for a given player is the sum of the independent $\mathcal{X}^2$ statistics across all his opponents with degrees of freedom equal to the sum of the degrees of freedom for each pair test. Additionally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

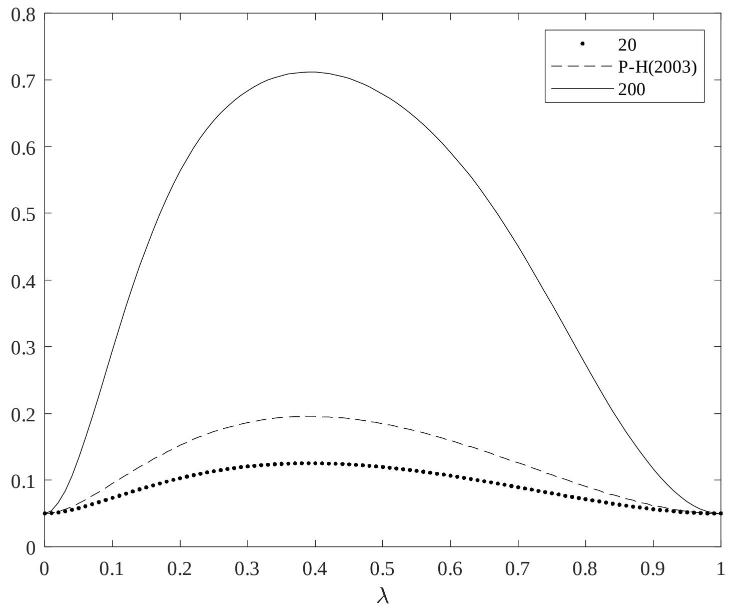
2.10 Graphs Chapter 2

Figure 2.1: Correlation of a "Heterogenous" Player



Notes: This graph displays the correlation between the actions of a supposedly homogeneous player with its scoring probability as a function of the fraction of observations from the first pair (λ).

Figure 2.2: Rejection Probabilities



Notes: This graph displays the rejection rate at the 5% nominal level of the non-central chi-square for 20, 38 (Palacios-Huerta (2003) median sample size) and 200 observations as a function of the fraction of observation from the first pair (λ).

Figure 2.3: p-value Plots of Tests for Independence with 2 Actions Under Heterogeneity

Figure 2.3.a: LPM LM

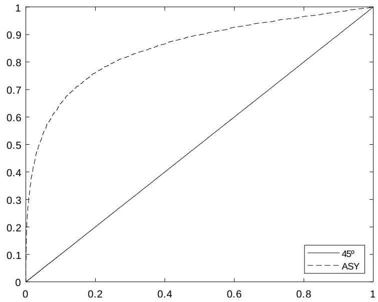


Figure 2.3.b: LPM LR

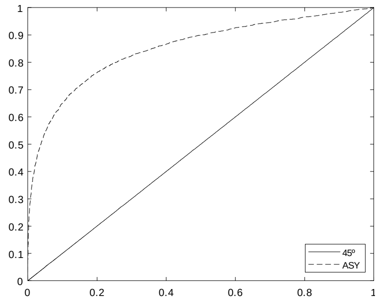


Figure 2.3.c: LPM F-test

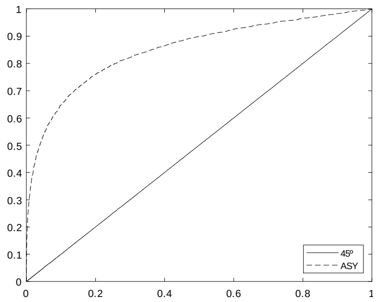


Figure 2.3.d: LPM Wald

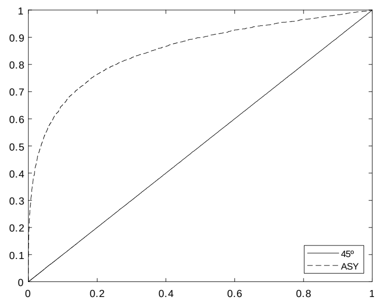


Figure 2.3.e: Logit LR

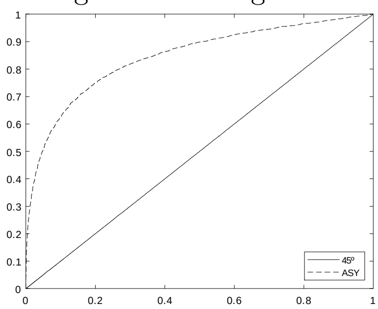
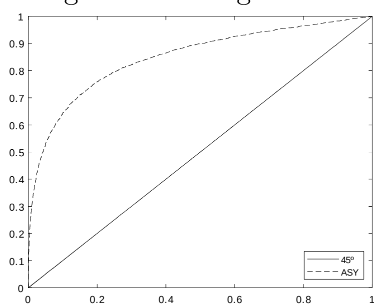


Figure 2.3.f: Logit Wald



Notes: These graphs display the empirical distribution functions of the asymptotic p-values (dashed line) in the Monte Carlo simulations for 200 observations for the following test statistics: LM test in the multivariate regression, the F-test, LR and Wald test in the multivariate regression, and the Wald and LR tests in the multinomial logit model (see Davidson and MacKinnon (1998)).

Figure 2.4: p-value Plots of Test of Equal Scoring Probabilities with 3 Actions

Figure 2.4.a: LPM LM Test

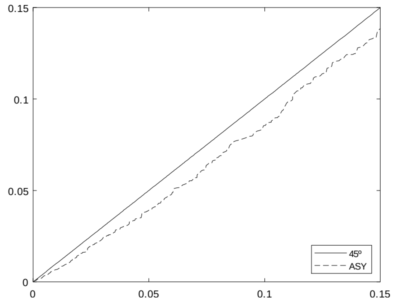


Figure 2.4.b: LPM LR Test

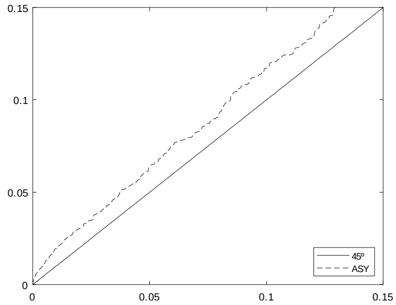


Figure 2.4.c: LPM F-Test

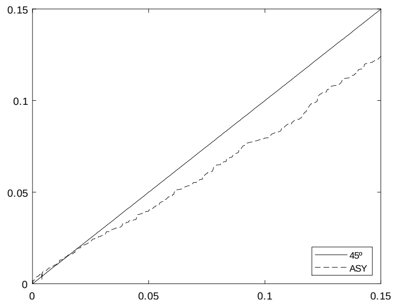


Figure 2.4.d: LPM Wald Test

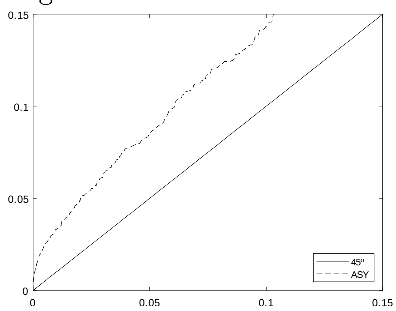


Figure 2.4.e: LPM Het. Rob. Wald Test

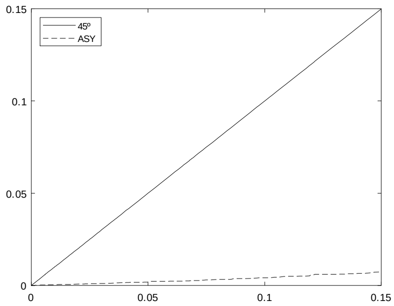


Figure 2.4.f: Logit LR Test

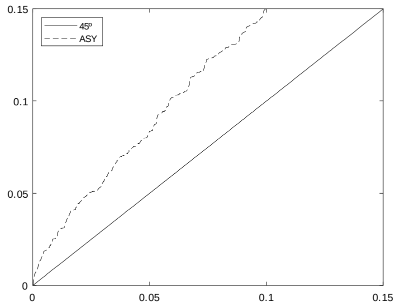
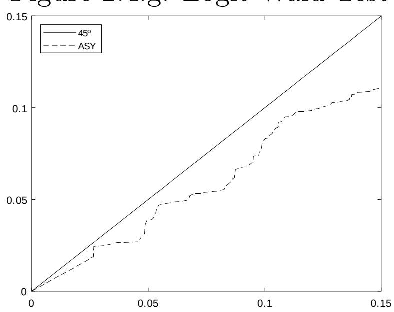


Figure 2.4.g: Logit Wald Test



Notes: These graphs display the empirical distribution functions of the asymptotic p-values (dashed line) in the Monte Carlo simulations for 200 observations for the following test statistics: LM test in the multivariate regression, the F-test, LR and Wald test in the multivariate regression, Wald's heteroskedasticity robust version, and the Wald and LR tests in the multinomial logit model (see Davidson and MacKinnon (1998)).

Figure 2.5: p-value Plots of Serial Independence Tests with 3 Actions

Figure 2.5.a: LPM LM Test

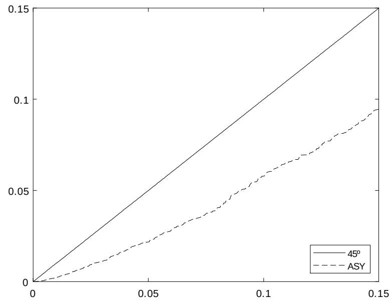


Figure 2.5.b: LPM Pillai Test

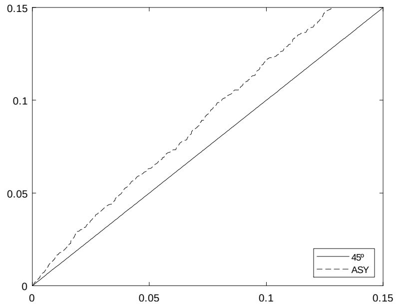


Figure 2.5.c: LPM LR Test

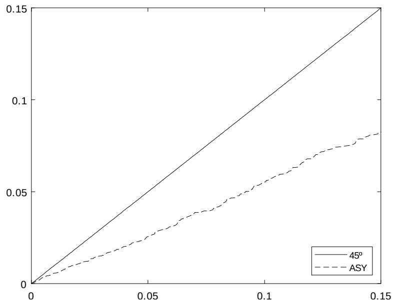


Figure 2.5.d: LPM Wilks' Test

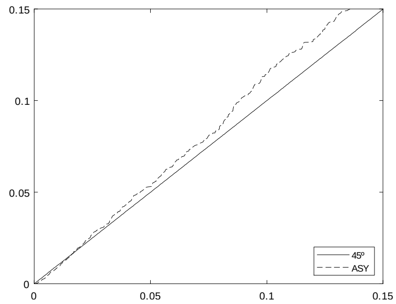


Figure 2.5.e: LPM Wald Test

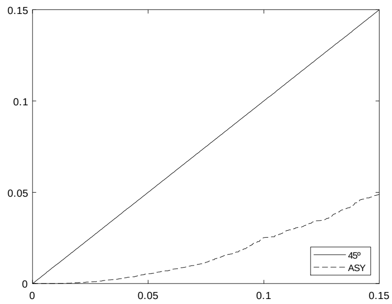


Figure 2.5.f: LPM L-H Test

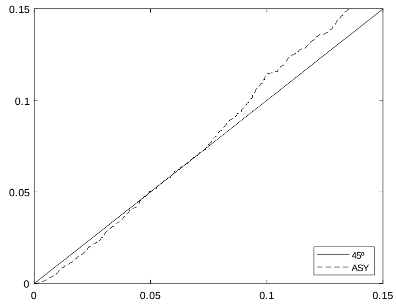


Figure 2.5.g: Logit LR Test

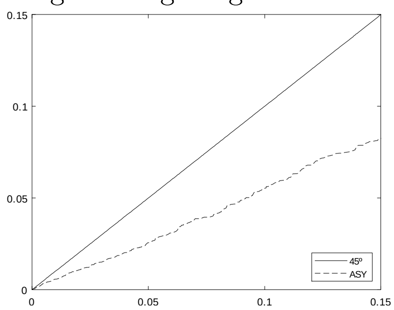
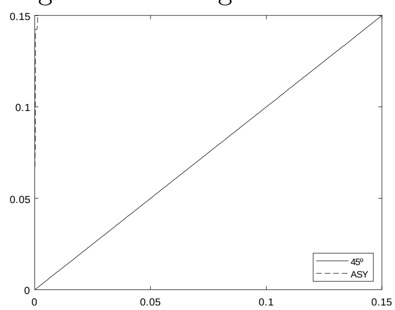


Figure 2.5.h: Logit Wald Test



Notes: These graphs display the empirical distribution functions of the asymptotic p-values (dashed line) in the Monte Carlo simulations for 200 observations for the following test statistics: LM test in the multivariate regression, Pillai trace, LR, Wilks' lambda, Wald and Lawley's-Hotelling trace test in the multivariate regression, and the Wald and LR tests in the multinomial logit model (see Davidson and MacKinnon (1998)).

Figure 2.6: p-value Plots of Tests for Action Independence with 3 Actions

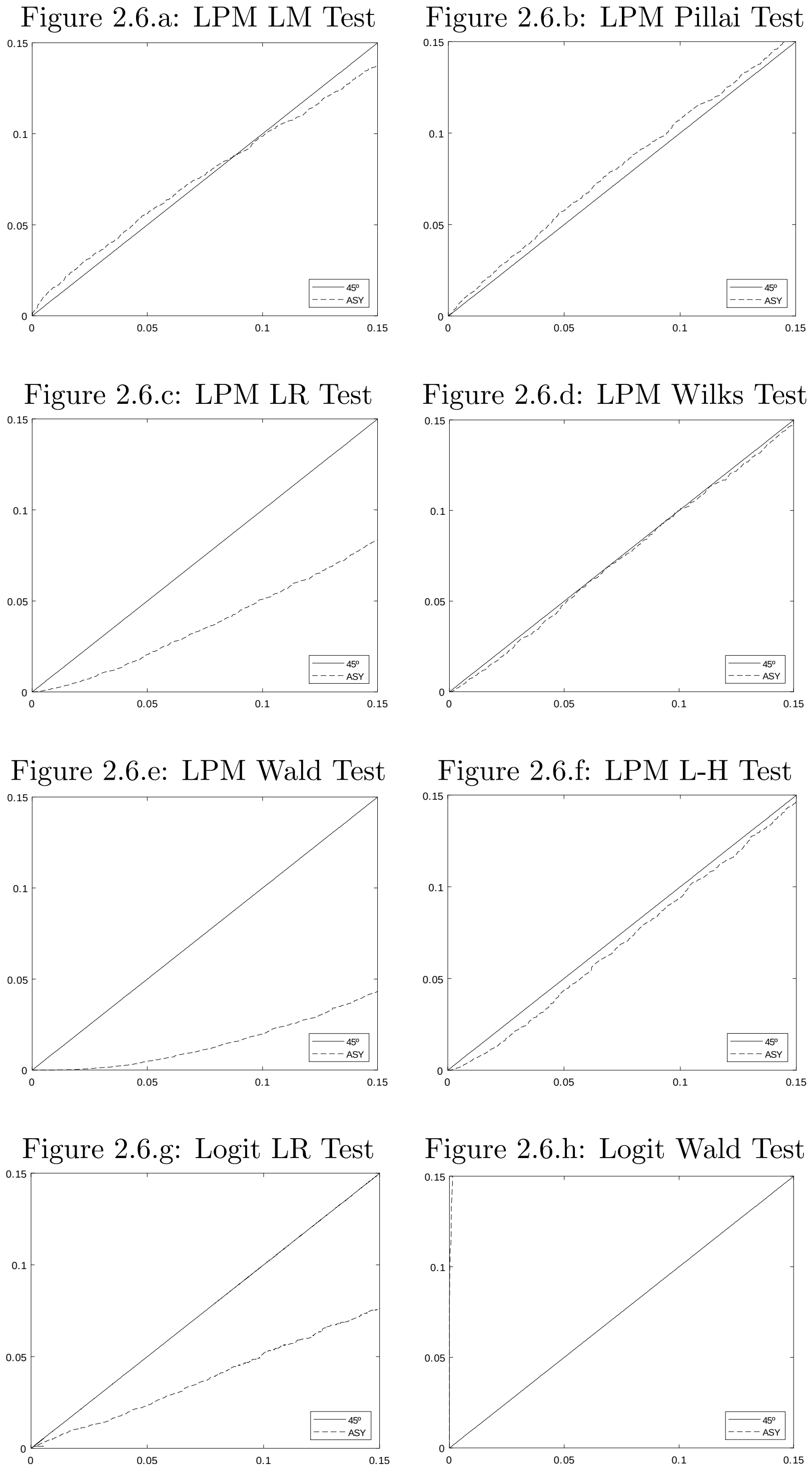


Figure 2.6.a: LPM LM Test

Figure 2.6.b: LPM Pillai Test

Figure 2.6.c: LPM LR Test

Figure 2.6.d: LPM Wilks Test

Figure 2.6.e: LPM Wald Test

Figure 2.6.f: LPM L-H Test

Figure 2.6.g: Logit LR Test

Figure 2.6.h: Logit Wald Test

Notes: These graphs display the empirical distribution functions of the asymptotic p-values (dashed line) in the Monte Carlo simulations for 200 observations for the following test statistics: LM test in the multivariate regression, Pillai trace, LR, Wilks' lambda, Wald and Lawley's-Hotelling trace test in the multivariate regression, and the Wald and LR tests in the multinomial logit model (see Davidson and MacKinnon (1998)).

Chapter 3

Tests For Independence Between Categorical Variables

3.1 Introduction

Economic theories are usually confronted with data to assess their validity. This is often done by deriving hypotheses implied by a theory and testing them with econometric procedures. In many important cases, such as testing the implications of mixed strategy equilibrium or the efficient market hypothesis, those hypotheses imply the independence between two categorical variables. There are multiple procedures in the literature one can use to conduct such tests, which leads to the crucial question of which approach to use for testing independence between categorical variables. Anatolyev and Kosenok (2009) showed the asymptotic equivalence between Pearson's independence test and the usual Wald test in a multivariate Linear Probability Model (LPM). However, this equivalence does not prevent that those tests lead to different conclusions in finite samples. In fact, it is even possible that researchers could report contradictory results with the same dataset. In addition, econometricians often prefer probit or logit models instead of LPMs.

The first contribution of this chapter is to prove the numerical equivalence for general categorical variables between (i) Pearson's independence test in a contingency table, (ii) the LM test in several popular regression models: the multivariate LPM, the conditional and unconditional multinomial model, the multinomial logit and probit models; and (iii) the corresponding J-test statistic for overidentifying restrictions in the Generalized Methods of Moments (GMM). Therefore, different researchers using different econometric procedures will reach the same conclusions if they use any of those tests.

Additionally, I show that the asymptotically equivalent LR tests of independence in the conditional and unconditional multinomial model, multinomial logit and probit models numerically coincide too. Finally, I prove that the heteroskedasticity-robust Wald tests in the multivariate LPM and GMM are numerically identical to the Wald test in the conditional multinomial model.

Given that the LM test is numerically the same in all those models, all the other independence tests will also be asymptotically equivalent. Therefore, the only reason why researchers might reach different conclusions in empirical applications is because they use either LR or Wald versions, not because they use different models. Table 3.1 summarizes the theoretical results.

Table 3.1: Numerical and Asymptotic Equivalence Results

Models\Tests	LM (Gradient)	LR (Distance)	Wald	Wald Robust	J-test
Multivariate LPM	$\bigcirc$	$\bigtriangledown$	$\bigtriangledown$	$\square$	-
Unconditional Multinomial Model	$\bigcirc$	$\bigtriangleup$	$\bigtriangledown$	$\bigtriangledown$	-
Conditional Multinomial Model	$\bigcirc$	$\bigtriangleup$	$\square$	$\bigtriangledown$	-
Multinomial Probit	$\bigcirc$	$\bigtriangleup$	$\bigtriangledown$	$\bigtriangledown$	-
Multinomial Logit	$\bigcirc$	$\bigtriangleup$	$\bigtriangledown$	$\bigtriangledown$	-
GMM	$\bigcirc$	$\bigcirc$	$\bigtriangledown$	$\square$	$\bigcirc$

Specifically, it presents by rows the different models that empirical researchers have employed to test independence between categorical variables, while each column contains the various tests that they have at their disposal for a given model. The symbol $\bigcirc$ corresponds to Pearson's test and all its numerically equivalent versions, while $\bigtriangleup$ and $\square$ represent the LR and Wald tests in the multinomial model. Finally, $\bigtriangledown$ stresses the asymptotic equivalence among them.

Another related contribution of this chapter is to show that all these equivalences also apply to tests of serial independence of a discrete Markov chain, which can be regarded as a time series analog to the multinomial model.

A real life example of independence between categorical variables arises in soccer penalty kicks, which capture the theoretical setting of a two-person zero-sum game with no pure strategy Nash equilibria extremely well due to the clarity of the rules and the detailed structure of the simultaneous one-shot play. As is well known, in games with no pure strategy Nash equilibria, one fundamental theoretical implication is that the probability of winning should be the same regardless of the strategy chosen. Effectively, this requires independence between a dummy variable which indicates winning and a categorical variable that describes the strategies of the player. In addition, when the game is finite with a unique Nash equilibrium, the only subgame perfect equilibrium stipulates to play the same Nash equilibrium in every period (see chapter 14 of Osborne (2003) for more details). As a result, a second implication of the theory with a temporal dimension is that the actions of the players should be serially independent.

In the existing empirical literature, different researchers have used different econometric procedures to test the independence implications mentioned above. For example, Palacios-Huerta (2003) tested the first independence implication by means of a contingency table, while

Chiappori et al (2002) used Wald tests in a LPM. A third possibility is to test that the winning probabilities implied by a probit or logit model do not depend on the action taken by the players, as in Brown and Rosenthal (1990), who relied on the LR test instead. Similarly, for the second implication, one could use Wilks' lambda, Pillai trace or the Lawley-Hotelling trace tests frequently employed in multivariate analysis of variance (see Stewart (1995) for more details), as well as dynamic multinomial probit or logit models. In contrast, I eliminate the possibility of obtaining conflicting empirical conclusions by using my numerical equivalence results to assess if professional soccer players, who are among the highest paid sportsmen, satisfy these two independence implications using penalty kicks in actual professional soccer games. Specifically, I collected a dataset of 549 penalty kicks that include the names of the teams and players involved for each penalty kick, the choices taken and the outcome of the kick. Given that soccer players are experts at their game, my dataset provides a notable advantage over lab experiments because it is virtually impossible for lab individuals to be proficient at unfamiliar games in a limited timeframe. Empirically, I find that the behavior of some players is inconsistent with the implications of mixed strategy equilibrium, in the sense that winning probabilities are not identical across strategies. In contrast, I find that the second testable implication (players' actions are serially independent) holds for all the players. The first result differs from the existing evidence on mixed strategies in professional sports (e.g., Walker and Wooders (2002), Chiappori et al (2002), Palacios-Huerta (2003), Hsu et al (2007)), while the second result is consistent with the literature.

Sports provide an ideal setting to empirically test economic theories because the players are experienced professionals and the stakes are high (see e.g. Garicano et al (2005), Miklós-Thal and Ulrich (2016) and the references therein). Nevertheless, in other economic situations the theoretical predictions of a model also imply independence between categorical variables, so my analysis applies rather more generally. Important examples include tests of strict exogeneity of the movements in the price of a firm's product for the observed excess demand/supply (Bouissou et al (1986)), testing the efficient market hypothesis (Pesaran and Timmermann (1994)) and testing independence between donating blood and the levels of monetary compensation (Mellström and Johannesson (2008)). Details of how these tests can be mapped to my setting can be found in section 3.3.2.

The rest of the chapter is organized as follows. Section 3.2 explains the different econometric methods. In section 3.3, I discuss several empirical applications of independence tests. Section 3.4 presents my numerical equivalence results as well as their time series extensions, while section 3.5 contains the empirical results. This is followed by the conclusions.

3.2 Econometric Methodology

There are several approaches in the literature that one can use to test for independence between two categorical variables. Before explaining the exhaustive list of different econometric procedures, I briefly define the notation used across the analysis. Let x be a $K \times 1$ categorical variable that takes values $(A_1, \cdots, A_K)$, where $A_1, \ldots, A_K$ are K exhaustive and mutually exclusive dummy variables which fully characterize the categorical variable. Similarly, let $\tilde{y}$ be another $H \times 1$ categorical variable that takes values $(B_1, \ldots, B_H)$. Both A_k and B_h, for $k = 1, \ldots, K$ and $h = 1, \ldots, H$, are dummy variables equal to 1 if its corresponding categorical value is equal to its k^{th} or h^{th} value, respectively.

A contingency table summarizes the sample information as follows:

$\tilde{y} \backslash x$	A_1	$\cdots$	A_K	Sum
B_1	n_{11}	$\cdots$	n_{1K}	$n_{1\diamond}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
B_H	n_{H1}	$\cdots$	n_{HK}	$n_{H\diamond}$
Sum	n_{*1}	$\cdots$	n_{*K}	n

where n_{hk}, for $h = 1, \ldots, H$ and $k = 1, \ldots, K$, denotes the observed joint frequency; for example, n_{12} is the number of times that B_1 and A_2 are simultaneously 1 in the sample. Also, $n_{h\diamond} = \sum_{k=1}^{K} n_{hk}$ denotes the number of times that B_h is 1, $n_{*k} = \sum_{h=1}^{H} n_{hk}$ the number of times A_k is 1 and $n = \sum_{k=1}^{K} n_{*k} = \sum_{h=1}^{H} n_{h\diamond}$ the total number of observations.

3.2.1 *Pearson's contingency test*

This is the original and best known test for independence, which is given by:

$$Pearson = \sum_{k=1}^{K} \sum_{h=1}^{H} [n_{hk} - (n_{*k} n_{h\diamond}/n)]^2 \, (n/n_{*k} n_{h\diamond}) \,. \tag{3.1}$$

Under the null hypothesis, this statistic follows a $\mathcal{X}^2$ distribution with $(H-1) \times (K-1)$ degrees of freedom in large samples under appropriate regularity conditions (see Mood et al (1974)).[1]

[1] Specifically, in addition to random sampling, it requires all joint frequencies π_{hk}, $h = 1, ..., H$ and $k = 1, ..., K$, to be strictly positive and fixed, so that the observed joint frequencies n_{hk} will increase asymptotically at the

Unlike most other statistics, the $\mathcal{X}^2$ statistic in (3.1) has an easy to interpret expression, which provides information on exactly which estimated joint frequencies account for its value.

It is worth mentioning that if $n_{h\diamond}$ and n_{*k}, for $h = 1, \ldots, H$ and $k = 1, \ldots, K$, were fixed, (3.1) would then become Fisher's (1922) exact test (see section 3.7.4).

3.2.2 *Multivariate regression*

This is a technique that combines several regression equations with the same regressors, one for each dependent variable. Multivariate regression is useful here because one can write the relationship between the dummy variables B_h and A_k for $h = 1, \ldots, H$ and $k = 1, \ldots, K$ as the following multivariate LPM:

$$\left.\begin{array}{c} B_{1i} = \delta_{11} A_{1i} + \cdots + \delta_{1K} A_{Ki} + u_{1i} \\ \vdots \\ B_{Hi} = \delta_{H1} A_{1i} + \cdots + \delta_{HK} A_{Ki} + u_{Hi} \end{array}\right\}. \tag{3.2}$$

Given that both regressors and regressands are dummy variables, the coefficients of the explanatory variables are the probability of the different values of the multinomial variable $\tilde{y}$ given the other multinomial variable x. For instance,

$$\delta_{hk} = E(B_h | A_1 = 0, \ldots, A_k = 1, \ldots, A_K = 0) = \Pr(B_h = 1 | A_1 = 0, \ldots, A_k = 1, \ldots, A_K = 0).$$

Hence, the sum of δ_{h1} for $h = 1, \ldots, H$ is equal to 1 for all the columns in the matrix of regression coefficients. Therefore, the coefficients in the equation for B_{Hi} can be obtained from the other $H - 1$ equations because $B_{Hi} = 1 - \sum_{h=1}^{H-1} B_{hi}$. For that reason, I can cross out the last equation without loss of generality to avoid a singular covariance matrix (see Judge et al (1985) chapter 12, section 5 for more details).

Define $B'_h = (B_{h1}, \cdots, B_{hn})$, $u'_h = (u_{h1}, \cdots, u_{hn})$ and $\delta_h = (\delta_{h1}, \cdots, \delta_{hK})$ for $h = 1, \ldots, H-1$.

Also, define the matrices

$$X = \begin{pmatrix} A_{11} & \cdots & A_{K1} \\ \vdots & \ddots & \vdots \\ A_{1n} & \cdots & A_{Kn} \end{pmatrix} \quad \text{and} \quad Y = \begin{pmatrix} B_{11} & \cdots & B_{H-1,1} \\ \vdots & \ddots & \vdots \\ B_{1n} & \cdots & B_{H-1,n} \end{pmatrix}.$$

same rate as the sample size n.

Finally, define the matrix of regression coefficients

$$\Pi = \begin{pmatrix} \delta_{11} & \cdots & \delta_{1K} \\ \vdots & \ddots & \vdots \\ \delta_{H-1,1} & \cdots & \delta_{H-1,K} \end{pmatrix},$$

with $\delta = vec\,(\Pi')$ and $\Sigma_U = E(u_i u_i') = Var\,(u_i)$. In this way, the multivariate regression model in (3.2) can be written as $Y = X\Pi' + u$.

Under the assumptions of the classical regression model, the parameters of the model can be efficiently estimated by OLS equation by equation. The reason is that $\hat{\delta}^{GLS} = \hat{\delta}^{OLS}$ because the regressors in all the $H-1$ equations are identical.

The OLS estimator of the parameters of the h^{th} equation is $\hat{\delta}_h^{OLS} = (X'X)^{-1}X'B_h$, where $X'X = diag(n_{*1},\ldots,n_{*K})$ and $X'B_h = (n_{h1},\ldots,n_{hK})'$ because $\sum_{i=1}^{n} A_{ki} = n_{*k}$ and $\sum_{i=1}^{n} A_{ki}B_{hi} = n_{hk}$ for $k = 1,\ldots,K$ and $h = 1,\ldots,H-1$.

This yields $\hat{\delta}_h^{OLS} = (n_{h1}/n_{*1},\ldots,n_{hK}/n_{*K})'$, so equation by equation OLS gives the natural estimator of δ_{hk}. Thus, the estimated probabilities are always non-negative and add up to 1, which avoids a common criticism of LPMs (see Wooldridge (2002)).

The null hypothesis of independence implies that $\delta_{h1} = \cdots = \delta_{hK} = \delta_h$ for $h = 1,\ldots,H-1$, so that the conditional and unconditional probabilities of $B_h = 1$ are the same. The restricted model can be estimated efficiently by OLS equation by equation because GLS is once more equal to OLS. The restricted OLS estimators are trivially $\tilde{\delta}_h^{OLS} = n_{h\diamond}/n$, for $h = 1,\ldots,H-1$.

The multivariate LPM has one potentially important disadvantage. Under the alternative, it violates the homoskedasticity assumption because the conditional variance of the error term u will change depending on the values of the explanatory variables rather than being the assumed constant matrix Σ_U (see Wooldridge (2002)). However, Σ_R, which is the covariance matrix of u given the dummy regressors under the null hypothesis of independence, is constant. This implies that the homoskedasticity assumption holds under the null, and hence all the usual regression tests are asymptotically valid.

The three classical multivariate regression tests are the $Wald$, LR and LM. Note that for any dataset, the relationship between these tests in a multivariate regression model is $Wald \geq LR \geq LM$ despite being asymptotically equivalent (see Berndt and Savin (1977) and Engle (1983) for more details). Moreover, they are monotonic transformations of the regression F-test in the $H = 2$ case only but not in general (see section 3.7.5 for the case of $H = 3$).

These three tests can be easily transformed into the Pillai trace, Wilks' lambda and Lawley-Hotelling trace tests used in multivariate analysis of variance (see Stewart (1995) for more details). More precisely, the Pillai trace test can be written as $V = n^{-1}LM$, while Wilks' lambda is $\Lambda = \exp\left(-n^{-1}LR\right)$ and the Lawley-Hotelling trace test is $LH = n^{-1}Wald$.

Finally, I consider a robust test which is still valid if the homoskedasticity assumption is violated. Specifically, I use the heteroskedasticity-robust Wald test in a multivariate regression, which I derive in section 3.7.3 using the results in Hansen (1982).

3.2.3 *Conditional multinomial model*

The LPM is usually regarded as a linear approximation to the true conditional probabilities. For that reason, define $P_{hk} = \Pr(B_h = 1|A_1 = 0, \ldots, A_k = 1, \ldots, A_K = 0)$ for $k = 1, \ldots, K$ and $h = 1, \ldots, H-1$, so that the joint probability of $B_h = 1$ and $A_k = 1$ is $\pi_{hk} = P_{hk} \times \pi_{*k}$, where $\pi_{*k} = \Pr(A_k = 1)$. Hence, the log-likelihood of the sample becomes:

$$\begin{aligned}\ln \mathcal{L} = \sum\nolimits_{k=1}^{K} \left[\left(n_{*k} - \sum\nolimits_{h=1}^{H-1} n_{hk} \right) \ln \left(1 - \sum\nolimits_{h=1}^{H-1} P_{hk} \right) + \sum\nolimits_{h=1}^{H-1} \left(n_{hk} \ln P_{hk} \right) \right] \\ + n_{*K} \ln \left(1 - \sum\nolimits_{k=1}^{K-1} \pi_{*k} \right) + \sum\nolimits_{k=1}^{K-1} n_{*k} \ln \pi_{*k} \qquad (3.3)\end{aligned}$$

because the number of times $A_k = 1$ and $A_k B_h = 1$ is n_{*k} and n_{hk}, respectively.

Maximizing the log-likelihood with respect to P_{hk} and π_{*k} yields $\hat{P}_{hk} = n_{hk}/n_{*k}$ and $\hat{\pi}_{*k} = n_{*k}/n_{*K}$, so that $\hat{P}_{hk} = \hat{\delta}_{hk}$ for $h = 1, \ldots, H-1$ and $k = 1, \ldots, K$.

Under the null, which states that $P_{hk} = P_{h\diamond}$ for $k = 1, \ldots, K$ and $h = 1, \ldots, H-1$, then

$$\begin{aligned}\ln \mathcal{L} = \left(n - \sum\nolimits_{h=1}^{H-1} n_{h\diamond} \right) \ln \left(1 - \sum\nolimits_{h=1}^{H-1} P_{h\diamond} \right) + \sum\nolimits_{h=1}^{H-1} n_{h\diamond} \ln P_{h\diamond} \\ + n_{*K} \ln \left(1 - \sum\nolimits_{k=1}^{K-1} \pi_{*k} \right) + \sum\nolimits_{k=1}^{K-1} n_{*k} \ln \pi_{*k}\end{aligned}$$

which yields $\tilde{P}_{h\diamond} = n_{h\diamond}/n$ and $\tilde{\pi}_{*k} = n_{*k}/n_{*K}$, so that $\tilde{P}_{h\diamond} = \tilde{\delta}_h$ for $h = 1, \ldots, H-1$. Note that $\hat{\pi}_{*k}$ will coincide under the null and alternative, so I can ignore these parameters.

As in the multivariate regression, I consider the LM, LR and $Wald$ tests (see section 3.7.1). It is worth mentioning that the information matrix evaluated under the null is block diagonal between P_{hk} and π_{*k}, which simplifies those expressions.

3.2.4 *Unconditional multinomial model*

Following the traditional treatment in Mood et al (1974) section 3.5.4, I can write the joint likelihood in terms of the joint probabilities $\pi_{hk} = \Pr(B_h = 1; A_k = 1)$ rather than P_{hk} and π_{*k}, with $\pi_{hk} = P_{hk} \times \pi_{*k}$. The null hypothesis states that there is independence between $\tilde{y}$ and x, so the joint probability should be the product of their marginal probabilities. In other words:

$$H_0 : \pi_{hk} = \pi_{h\diamond} \times \pi_{*k} \ , \ h = 1, \ldots, H \text{ and } k = 1, \ldots, K.$$

For this reason, it is convenient to write the joint probabilities under the alternative as the product of two sets of parameters: (i) $\pi_{h\diamond}$, for $h = 1, \ldots, H-1$, and π_{*k}, for $k = 1, \ldots, K-1$, which denote the parameters of the marginal probability distribution for $\tilde{y}$ and x respectively, and (ii) $(K-1) \times (H-1)$ additional parameters ϑ which should be 0 under the null (see again Mood et al (1974)).

In particular, in the 3×3 case I can write the joint probabilities as

$$\left.\begin{array}{l}
\pi_{11} = \pi_{1\diamond}\pi_{*1} + \vartheta_{11} \ , \ \pi_{12} = \pi_{1\diamond}\pi_{*2} + \vartheta_{12} \\
\pi_{13} = \pi_{1\diamond}(1 - \pi_{*1} - \pi_{*2}) - \vartheta_{11} - \vartheta_{12} \ , \ \pi_{21} = \pi_{2\diamond}\pi_{*1} + \vartheta_{21} \\
\pi_{22} = \pi_{2\diamond}\pi_{*2} + \vartheta_{22} \ , \ \pi_{23} = \pi_{2\diamond}(1 - \pi_{*1} - \pi_{*2}) - \vartheta_{21} - \vartheta_{22} \\
\pi_{31} = (1 - \pi_{1\diamond} - \pi_{2\diamond})\pi_{*1} - \vartheta_{11} - \vartheta_{21} \ , \ \pi_{32} = (1 - \pi_{1\diamond} - \pi_{2\diamond})\pi_{*2} - \vartheta_{12} - \vartheta_{22} \\
\pi_{33} = (1 - \pi_{1\diamond} - \pi_{2\diamond})(1 - \pi_{*1} - \pi_{*2}) + \vartheta_{11} + \vartheta_{12} + \vartheta_{21} + \vartheta_{22}
\end{array}\right\} ,$$

where the additional free parameters ϑ_{11}, ϑ_{12}, ϑ_{21} and ϑ_{22} become 0 under the null.[2]

The same procedure can be applied to the general $H \times K$ case. Analogous derivations to the ones in the previous section show that the estimators of the marginal probabilities are the same under the null and the alternative, and that the information matrix evaluated under the null is block diagonal between $\pi_{h\diamond}$, π_{*k} and the ϑ's.

3.2.5 *Multinomial probit model*

Following section 27.3 of Ruud (2000), consider the following "random utilities" model:

$$\left.\begin{array}{c}
B^*_{1i} = \alpha_{11}A_{1i} + \cdots + \alpha_{1K}A_{Ki} + \varepsilon_{1i} \\
\vdots \\
B^*_{Hi} = \alpha_{H1}A_{1i} + \cdots + \alpha_{HK}A_{Ki} + \varepsilon_{Hi}
\end{array}\right\} ,$$

[2] In the 2×2 case, the joint probabilities will be π_{11}, π_{12}, π_{21} and π_{22}, which depend on $\pi_{1\diamond}$, π_{*1} and only one other parameter, namely $\vartheta = \pi_{12} - \pi_{*1}\pi_{2\diamond}$, which should be 0 under the null.

where $\varepsilon_h|x \sim i.i.d.\ N(0,\omega)$, with x being defined at the beginning of section 3.2.

Let the observation rule be $B_{hi} = 1\left\{B^*_{hi} = \max_{j=1,\dots,H} B^*_{ji}\right\}$, where $1\{\}$ is the indicator function, so that $B_{hi} = 1$ if h is the preferred choice. In other words, one chooses the action h that maximizes one's utility. The conditional log-likelihood function is

$$L(\theta) = \sum\nolimits_{i=1}^{n} \sum\nolimits_{h=1}^{H} B_{hi} \ln \Pr(B_{hi} = 1|x_i),$$

where θ contains the model parameters α_h (slope coefficients) and ω (covariance matrix).[3] Under the null hypothesis, $\alpha_{h1} = \cdots = \alpha_{hK} = \alpha_h$, for $h = 1,\dots,H$.

This log-likelihood function coincides with the conditional component of the log-likelihood function of the multinomial model in (3.3) but expressed in terms of α's. However, the probabilities involve multiple normal integrals of dimension $H-1$.

3.2.6 *Multinomial logit model*

As is well known (see section 27.4 of Ruud (2000)), we will get the multinomial logit model if in the random utility model above, ε_{hi}, instead of being normal, is drawn from an *i.i.d.* extreme value distribution. Define the conditional probability matrix P as:

$$P = \begin{pmatrix} P_{11} & \cdots & P_{1K} \\ \vdots & \ddots & \vdots \\ P_{H1} & \cdots & P_{HK} \end{pmatrix}$$

with P_{hk} being the same as in the conditional multinomial model. This model ensures the non-negativity of P_{hk}, for all h,k, as well as the adding up constraint $\sum_{h=1}^{H} P_{hk} = 1$, by assuming the following functional form:

$$\left.\begin{array}{l} \Pr(B_h = 1 \mid A_1,\dots,A_K) = (1+D)^{-1} \exp\left(\sum_{k=1}^{K} \gamma_{hk} A_{ki}\right),\ h = 1,\dots,H-1 \\ \Pr(B_H = 1 \mid A_1,\dots,A_K) = (1+D)^{-1} \end{array}\right\},$$

where $D = \sum_{h=1}^{H-1} \exp\left(\sum_{k=1}^{K} \gamma_{hk} A_{ki}\right)$, and γ_{hk}, for $h = 1,\dots,H-1$ and $k = 1,\dots,K$, are the model parameters. Therefore, like in the multinomial probit, the multinomial logit is effectively a reparametrization of the matrix P which ensures non-negative probabilities that add up to 1 by columns.

This log-likelihood function is analogous to the conditional component of the likelihood in

[3] The independent probit model without dependence across ε_{hi}'s is the most flexible model that one can identify in this case.

(3.3) but expressed in terms of γ_{hk} instead of P_{hk}'s.

The null hypothesis of independence states that $H_0 : \gamma_{h1} = \ldots = \gamma_{hK} = \gamma_h$ for $h = 1, \ldots, H-1$.

3.2.7 *GMM*

Given that $P_{hk} = \Pr(B_h = 1|A_1 = 0, \ldots, A_k = 1, \ldots, A_K = 0)$, we can express all those parameters in terms of the following set of moment conditions

$$E\left[(y_i - \Pi x_i) \otimes x_i\right] = 0, \tag{3.4}$$

where $\otimes$ denotes Kronecker product and y is a categorical variable that takes values $(B_1, \ldots, B_{H-1})$, so that it coincides with the first $H-1$ elements of $\tilde{y}$. These moment conditions coincide with the normal equations, which are the first order conditions of the multivariate LPM, as well as with the scores of the conditional multinomial model. The null hypothesis of independence implies that $H_0 : \pi_{hk} = \pi_h$, for $k = 1, \ldots, K$ and $h = 1, \ldots, H-1$. Under H_1, Π is unrestricted while under H_0, $\Pi = \upsilon l_K'$, where l_K is a vector of K ones. Note that under H_0, I can write $\Pi'(\upsilon) = l_K \upsilon' I_{H-1}$, which implies that $\delta(\upsilon) = vec(\Pi'(\upsilon)) = (I_{H-1} \otimes l_K)\upsilon$.

The GMM estimator is defined as:

$$\hat{\upsilon} = \arg\min_{\upsilon} \left(\frac{1}{n}\sum\nolimits_{i=1}^{n}\{[y_i - \Pi(\upsilon)x_i] \otimes x_i\}\right)' \Upsilon^{-1} \left(\frac{1}{n}\sum\nolimits_{i=1}^{n}\{[y_i - \Pi(\upsilon)x_i] \otimes x_i\}\right),$$

where Υ is a symmetric positive definite $[K \times (H-1)] \times [K \times (H-1)]$ weight matrix.

With random sampling, the optimal GMM estimator is the one which minimizes the GMM criterion function when Υ is equal to $[\Sigma_R \otimes \sum_{i=1}^{n}(x_i x_i')]$, where Σ_R is the covariance matrix of the multivariate LPM under the null.

The J-test statistic for overidentifying restrictions is just the value of the GMM objective function evaluated at the efficient GMM estimator (see Hansen (1982) for more details). Algebraically, $J = n \times \bar{g}(\hat{\upsilon})'\Upsilon^{-1}\bar{g}(\hat{\upsilon})$, where $\bar{g}(\hat{\upsilon}) = n^{-1}\sum_{i=1}^{n}\{[y_i - \Pi(\hat{\upsilon})x_i] \otimes x_i\}$.

3.3 Practical Applications

3.3.1 *Applications to mixed strategies in soccer penalty kicks*

Next, I illustrate the two main implications of mixed strategy equilibrium for a game with two players and three actions with the most popular econometric methods in section 3.2. First, consider the null hypothesis of equal scoring probabilities:

LPM Let $B = 1$ if the penalty kick is scored. The LPM under the alternative is defined as $B = \delta_L A_L + \delta_C A_C + \delta_R A_R + u$, where A_k, for $k = L, C, R$, are dummy variables. For example, A_R takes the value 1 if the penalty is shot in that direction and 0 otherwise. As we saw in section 3.2, the regression coefficients are conditional scoring probabilities. I can estimate the model by OLS. In turn, the LPM under the null hypothesis states that $\delta_L = \delta_C = \delta_R = \delta$.

The LM test is $LM = nR^2 = n \times [1 - (\hat{\sigma}_U^2 / \hat{\sigma}_R^2)]$, where $\hat{\sigma}_U^2$ is the unrestricted variance estimator, $\hat{\sigma}_R^2$ the restricted variance estimator, n the number of observations and R^2 the R-squared of the regression (see Wooldridge (2002) chapter 8.3).

Logit model The Logit model for the penalty kick game is described as follows:

$$\left.\begin{aligned} &B = 1 \text{ if } B^* \geq 0 \\ &B^* = \gamma_L A_L + \gamma_C A_C + \gamma_R A_R + \varepsilon \end{aligned}\right\},$$

where ε is logistically distributed, so that $\Pr(B = 1|X) = D/(1 + D)$, where $D = \exp(\gamma_L A_L + \gamma_C A_C + \gamma_R A_R)$. Therefore, in this model

$$P_R = \Pr(B = 1|A_L = 0, A_C = 0, A_R = 1) = \exp(\gamma_R) / \left(1 + \exp(\gamma_R)\right),$$

so $\gamma_R = \ln[P_R/(1 - P_R)]$. The Logit model under the null states that $\gamma_L = \gamma_C = \gamma_R = \gamma$.

The LM test statistic is $LM = n \times \left[s(\widetilde{\gamma})' \mathcal{I}(\widetilde{\gamma})^{-1} s(\widetilde{\gamma})\right]$, where $s(\widetilde{\gamma})$ is the unrestricted gradient evaluated at the restricted estimators $\widetilde{\gamma}$ and $\mathcal{I}(\widetilde{\gamma})$ is the unrestricted information matrix evaluated at the restricted estimators (see section 3.7.1).

Now, I turn to the serial independence hypothesis:

Multivariate LPM The multivariate regression one can use to detect possible departures from serial independence is similar to a first-order vector autoregressive process (see Wooldridge (2002) chapter 18, Section 5 for more details). Specifically:

$$\left.\begin{aligned} A_{L,t} &= \delta_{LL} A_{L,t-1} + \delta_{CL} A_{C,t-1} + \delta_{RL} A_{R,t-1} + u_{Lt} \\ A_{C,t} &= \delta_{LC} A_{L,t-1} + \delta_{CC} A_{C,t-1} + \delta_{RC} A_{R,t-1} + u_{Ct} \\ A_{R,t} &= \delta_{LR} A_{L,t-1} + \delta_{CR} A_{C,t-1} + \delta_{RR} A_{R,t-1} + u_{Rt} \end{aligned}\right\},$$

where $A_{k,t}$, for $k = L, C, R$, are the dependent variables and $A_{k,t-1}$, for $k = L, C, R$, are lagged regressors. In this multivariate regression with three lagged explanatory variables, but no constant, the coefficients of the lagged variables are the probability of choosing a strategy at time t conditional on the previous action. These are sometimes called transition probabilities. For instance, δ_{RL} measures the probability of $A_{L,t}$ being equal to 1 given that $A_{R,t-1}$ is equal to 1, etc.

The null hypothesis of independence implies that the δ's have to be the same across rows. The formula for the LM test in this model is:

$$LM = (n-1) \times tr\left[\left(\frac{n-1}{n-2} \times \hat{\Sigma}_R^{-1}\right) \times \left(\frac{n-2}{n-1} \times \hat{\Sigma}_R - \frac{n-4}{n-1} \times \hat{\Sigma}_U\right)\right]$$

where $\hat{\Sigma}_R$ and $\hat{\Sigma}_U$ are the restricted and unrestricted estimates of the residual variance covariance matrices and n the number of observations (see section 3.7.1).

Multinomial logit model Let x_t denote the action chosen by the player at time t. The multinomial logit model for the penalty kick game is described as follows:

$$\left.\begin{array}{l} \Pr(x_t = A_L \mid x_{t-1}) = \exp(\gamma_{LL}A_{L_{t-1}} + \gamma_{LC}A_{C_{t-1}} + \gamma_{LR}A_{R_{t-1}}) / (1+D) \\ \Pr(x_t = A_C \mid x_{t-1}) = 1/(1+D) \\ \Pr(x_t = A_R \mid x_{t-1}) = \exp(\gamma_{RL}A_{L_{t-1}} + \gamma_{RC}A_{C_{t-1}} + \gamma_{RR}A_{R_{t-1}}) / (1+D) \end{array}\right\},$$

with

$$D = \exp(\gamma_{LL}A_{L_{t-1}} + \gamma_{LC}A_{C_{t-1}} + \gamma_{LR}A_{R_{t-1}}) + \exp(\gamma_{RL}A_{L_{t-1}} + \gamma_{RC}A_{C_{t-1}} + \gamma_{RR}A_{R_{t-1}}).$$

Under the null hypothesis, $\gamma_{LL} = \gamma_{LC} = \gamma_{LR} = \gamma_L$ and $\gamma_{RL} = \gamma_{RC} = \gamma_{RR} = \gamma_R$.

3.3.2 *Other empirical applications*

Price changes and excess supply Bouissou et al (1986) used a panel of French firms to investigate the relationship between movements in the price of their products and their observed excess demand/supply. Specifically, they used data across firms to test whether the sign of the price changes of a product over the last quarter could be regarded as strictly exogenous for the existence of excess demand or supply for that product.

The procedures that I explained in section 3.2 to test independence between categorical variables can be easily used here as follows. Let $B = 1$ if a firm's product exhibits excess supply and 0 otherwise. Similarly, let A_I, A_C and A_D denote three dummy variables indicating whether the price of that product has increased, remained constant or decreased in the last quarter. In this example, the LPM under the alternative would be defined as $B = \delta_I A_I + \delta_C A_C + \delta_D A_D + u$, so that $\delta_k = \Pr(B = 1, A_k = 1) / \Pr(A_k = 1)$, for $k = I, C, D$, represents the corresponding conditional probabilities.

Efficient market hypothesis Another interesting empirical example is Pesaran and Timmermann (1994), who were interested in testing the efficient market hypothesis in financial markets. They showed that lack of directional predictability of asset returns can be interpreted as stochastic independence between the sign of the actual returns and the sign of the predictions made by asset managers who want to time the market (see also Henriksson and Merton (1981) and Swanson and White (1997)).

Let B be a dummy variable that indicates if the realized value of the excess return on asset is positive or negative, and let A_P, A_Z and A_N denote three dummy variables indicating whether its forecast is positive, zero or negative. In this case, the LPM under the alternative would be defined as $B = \delta_P A_P + \delta_Z A_Z + \delta_N A_N + u$.

Blood donations Mellström and Johannesson (2008) conducted a field experiment with three different treatments to test whether the probability of becoming a blood donor is independent of the monetary compensation offered in each of those treatments.

Once again, let $B = 1$ if the subject agrees to become a blood donor and A_{NP}, A_P and A_{PC} denote three dummy variables indicating whether the subject receives no payment, a single monetary payment or a payment with a charity option, respectively. As expected, the LPM under the alternative would be defined as $B = \delta_{NP} A_{NP} + \delta_P A_P + \delta_{PC} A_{PC} + u$.

3.4 Numerical Equivalence Results

The main theoretical result in this chapter is that the LM tests in all the popular linear and non-linear regression models discussed in section 3.2 coincide with Pearson's test for independence as well as with the J-test statistic for overidentifying restrictions. The following proposition, which I prove in section 3.7.1, contains the precise result:

Proposition 3.1 *For general H and K, the Lagrange Multiplier test statistic for independence in a multivariate linear probability model, multinomial logit, multinomial probit and the conditional and unconditional multinomial models, computed using the information matrix, are numerically identical to Pearson's contingency table test statistic for independence and the J-test statistic for overidentifying restrictions in GMM. Additionally, the same numerical equivalence result holds if one exchanges regressors and regressands in all those models.*

This means that different researchers using different econometric procedures will reach exactly the same conclusions if they use LM tests. From the computational point of view, the easiest test is Pearson's statistic, which has the very simple closed-form expression in (3.1). In contrast, the multinomial logit and especially probit models should be avoided because they require numerical optimization and multiple integrals in the second case.

Another implication of Proposition 3.1 is that there will only be one finite sample distribution for all those different tests (see section 3.7.4 for more details). Additionally, the Monte Carlo experiments previously reported in the literature on Pearson's test also apply to all the other different tests, so they could be combined in a meta study.

Proposition 1 also says that if we exchange x and $\tilde{y}$ so that x now takes values $B_1, \ldots, B_H$ and $\tilde{y}$ takes values $A_1, \ldots, A_K$, then the corresponding test statistics will not change. While this is obvious for the Pearson's test (3.1) because the contingency table in section 3.2 will simply be flipped, it is far from obvious for all the other models.

For example, one obtains numerically the exact same LM statistic if one regresses $\tilde{y}_i$ on x_i or x_i on $\tilde{y}_i$ in the multivariate LPM. Similarly, imposing independence on $\Pr(B_h = 1 \mid A_1, \ldots, A_K)$ for all h yields the same LM statistic in a conditional multinomial model as imposing it on $\Pr(A_k = 1 \mid B_1, \ldots, B_H)$ for all k. As an illustration, one could test the independence of the kicker's action from the outcome of the kick and obtain exactly the same result.

It is worth mentioning that the numerical result in Proposition 3.1 is substantially different from the famous numerical inequality in Berndt and Savin (1977), which implies that $Wald \geq LR \geq LM$ in the multivariate linear probability model.[4] In contrast, I show that the LM test is numerically identical across models.

Additionally, four of the models in section 3.2 are essentially the same. Specifically, the log-likelihood function under the null and alternative of the multinomial logit and probit models are analogous to the corresponding conditional component of the log-likelihood of the multinomial model. In addition, the unconditional multinomial model can be regarded as an alternative reparametrization of the joint probabilities. Therefore, I also prove in section 3.7.2 the following equality:

Proposition 3.2 *For general H and K, the Likelihood Ratio test statistic for the null hypothesis of independence in the multinomial logit, multinomial probit and the conditional and unconditional multinomial models are numerically identical.*

This means that even though one can use any of those four different econometric models, the conclusions and implications will also be the same if one uses LR tests.

[4] See Dastoor (2001) for alternative inequalities in other models.

Although the Wald tests in all the models in section 3.2 will generally differ, the numerical equivalence between the OLS estimator of the regression coefficients in the multivariate LPM, the ML estimators of the conditional probabilities, and the unrestricted GMM estimators suggest a close relationship. It turns out that the crucial difference is the homoskedasticity assumption in the standard Wald test of the multivariate LPM. Specifically, if one decided to carry out a robust test which would remain valid when the homoskedasticity assumption is violated, the following numerical equality, which I prove in section 3.7.3, will hold:

Proposition 3.3 *For general H and K, the heteroskedasticity-robust Wald test statistic for independence in the multivariate LPM and GMM is numerically identical to the Wald test statistic of the conditional multinomial model.*

This implies that any of those tests will yield the same results and implications.

Furthermore, Table 3.1 in the introduction uses symbols to highlight the numerical equivalence results in Propositions 3.1-3. It also indicates that all the other remaining independence tests are not numerically identical. For example, $LR_{LPM} \neq LR_{Multinomial}$ because the true conditional distribution of the LPM is not normal, so the (pseudo) likelihood function of the multivariate regression is different from the likelihood of the multinomial model even under the null (see section 3.7.1). Similarly, the Wald test of the multinomial logit model is different from the multinomial version in the conditional multinomial model, and the same applies to the multinomial probit model because Wald tests are not invariant to non-linear transformations of the restrictions, despite having the same log-likelihood functions under the null and alternative.

Nevertheless, all the tests in Table 3.1 are asymptotically equivalent within each model, as shown in section 17.3 of Ruud (2000). Given that the LM test is numerically equivalent in all those models, all the other tests in Table 3.1 will also be asymptotically equivalent. Therefore, the only reason why researchers might reach different conclusions in empirical applications is precisely because they use Wald or LR versions rather than LM tests.

3.4.1 Serial independence tests for Markov chains

Next, I extend the numerical equivalence results in Propositions 3.1-3 in *i.i.d* contexts to serial independence tests for discrete Markov chains.

Let us summarize the K strategies for each player $(A_1, \ldots, A_K)$ at time t by means of the vector x_t, which has the Markov property if for all $k \geq 1$ and all t

$$\Pr(x_{t+1}|x_t, x_{t-1}, x_{t-2}, \ldots x_{t-k}) = \Pr(x_{t+1}|x_t).$$

In this context, the Markov chain is fully characterized by the $K \times K$ transition matrix

$$P = \begin{pmatrix} P_{11} & \cdots & P_{1K} \\ \vdots & \ddots & \vdots \\ P_{K1} & \cdots & P_{KK} \end{pmatrix},$$

where $P_{hk} = \Pr(x_{t+1} = x_h | x_t = x_k)$ are the one step transition probabilities with states $k = 1, \ldots, K$, where $P_{Kk} = 1 - \sum_{h=1}^{K-1} P_{hk}$, for all k and $h = 1, \ldots, K-1$.

Let n_{hk} be the number of times that there occurs a one period transition from state k to state h, with $A_{k1} = 1$ if the first observation belongs to state k and zero otherwise. The likelihood function of the Markov chain can be written as:

$$L(\theta) = P(x_1) \prod_{h=1}^{H} \prod_{k=1}^{K} P_{hk}^{n_{hk}},$$

where $P(x_1) = \prod_{k=1}^{K} \pi_k^{A_{k1}}$ and $\theta = (P_{11}, \ldots, P_{KK})$ (see Lee et al (1968) for more details). Hence, the log-likelihood function is:

$$\mathcal{L}(\theta) = \sum_{k=1}^{K} \left[n_{Kk} \ln \left(1 - \sum_{h=1}^{K-1} P_{hk} \right) + \sum_{h=1}^{K-1} (n_{hk} \ln P_{hk}) \right]$$
$$+ A_{K1} \ln \left(1 - \sum_{h=1}^{K-1} \pi_k \right) + \sum_{k=1}^{K-1} A_{k1} \ln \pi_k.$$

Note that this log-likelihood function is different from (3.3) because the marginal model is based on a single observation while the conditional model is recursive.

In contrast, if the Markov chain is serially independent, the matrix P will be:

$$P = l_K \times \begin{pmatrix} \pi_1 & \cdots & \pi_{K-1} & 1 - \sum_{k=1}^{K-1} \pi_k \end{pmatrix}.$$

I can achieve this by imposing the null hypothesis $H_0 : P_{hk} = P_h$, for $k = 1, \ldots, K$ and $h = 1, \ldots, K-1$ because $P_K = 1 - \sum_{h=1}^{K-1} P_h$. I can then obtain the restricted estimators from the following log-likelihood:

$$\mathcal{L}(\phi) = \sum_{k=1}^{K} \left[\sum_{h=1}^{K-1} (n_{hk} \ln P_h) + n_{Kk} \ln \left(1 - \sum_{h=1}^{K-1} P_h \right) + n_{*k} \ln \pi_k \right],$$

where $\phi = (P_1, \ldots, P_{K-1})$ and $n_{h\#} = \sum_{k=1}^{K} n_{hk}$.

Serial independence can then be assessed by means of the usual Wald, LR and LM tests. Therefore, I can easily show that despite the apparent differences, the numerical equivalence results in Propositions 3.1-3 also apply to test for serial independence in Markov chains.

3.5 Empirical application to penalty kicks

Soccer is one of the most important sports in the world. In fact, professional soccer players are among the highest paid sportsmen. For instance, the 2018 average annual pay for Barcelona and Real Madrid players exceeded those of all the US baseball (MLB), football (NFL) and basketball (NBA) teams. The main purpose of this section is to check if soccer players really behave as game theory predicts in such a high stakes context.

It is well known that penalty kicks decide matches, qualifications for next rounds in tournaments and even titles. Therefore it is not surprising that soccer teams devote resources to analyze and improve strategies for their players. Although most players and coaches do not know this, penalty kicks are a relevant example of a two-person zero-sum game due to the clarity of the rules and the detailed structure of the simultaneous one-shot play.

In this section, I use the econometric methods previously described to test if the empirical results obtained by Chiappori et al (2002) and Palacios-Huerta (2003) are still valid. Given that their datasets are not publicly available, I construct a similar but more recent dataset which contains 549 penalty kicks.[5] Moreover, I have expanded the actions of the players to six for a presumably tougher test of the predictions of mixed strategy equilibrium. In my dataset, there are 12 kickers with more than 20 penalty kicks and another 11 kickers with at least 10 penalties. Similarly, there are 10 goalkeepers with more than 10 observations. The identities of goalies and kickers are shown in section 3.7.6.

The penalty kick data I have collected covers the period 2005-2015 from professional games in Spain, Italy, England and other European countries. The information comes from the following Spanish TV programs and internet pages: Estudio Estadio (TVE), GOL TV, Canal + Liga, El Día Después (Movistar Plus), Deportes Cuatro, As.com and Marca.com. These TV programs and internet pages systematically review the best games played during the weekend, including all penalty kicks that take place in those games.

The data include the names of the teams involved in the match, the date, the names of the kicker and goalkeeper for each penalty kick, the choices taken: Left down (LD), Left up (LU), Center down (CD), Center up (CU), Right down (RD) and Right up (RU), the time and score at the time of the penalty, the final score of the game, the foot used by the kicker (left or right)

[5] There is no reason to expect substantive differences in the datasets despite covering different time spans because the rules governing penalty kicks have been the same for decades and team managers have always been aware of their importance.

and the outcome of the kick (goal or miss). The following table offers a basic description of the data with three actions.

(Table 3.2)

In particular, it shows the relative proportions of different choices made by both kickers and goalkeepers (see section 2.4.3 for more details).

The strategy followed by goalkeepers coincides with that followed by kickers in 42.8% of all penalties in the dataset. Kickers do not usually kick to the center (7.1% of all kicks), whereas goalies remain in the middle less often (4.55%). The percentage of kicks where the actions of the players do not coincide is mostly divided between LR (26.78%) and RL (20.95%). A goal is scored in 86.34% of all penalty kicks. The scoring rate is over 90% when the kicker choice differs from the goalie, and it is just over 65% when it coincides.

3.5.1 *Test of equal scoring probabilities*

The first testable implication I check is whether the scoring probabilities for a player are identical across strategies.[6] To compare my results with the results obtained by Palacios-Huerta (2003), I initially consider only the actions he took into account (Left and Right). To do so, I eliminate C for all players.[7] The results of the tests are described in Table 3.3. The null hypothesis of equal scoring probabilities across those two strategies is rejected for only one goalkeeper at the 5% level. But since the binomial probability of one or more goalkeepers out of 10 rejecting the null hypothesis at the 5% level when the null is true is 0.401, I do not reject the null hypothesis for those players when treated as a group (see section 2.8.7 for additional details on multiple tests). This finding agrees with Palacios-Huerta (2003).

A remarkable result is a 0 test statistic for kicker 3 because his scoring probabilities are the same regardless of the strategy chosen, so he behaves perfectly according to the theory. Similarly, goalkeeper 4 had a test statistic of 0, but this is due to the fact that his scoring probabilities are 100% regardless of the strategy chosen. From the point of view of his team, this result implies that he is not a very good performer when it comes to saving a penalty kick. Ironically, though, he chooses his strategies "optimally" from a game theoretical point of view.

Given that the empirical description in Table 3.2 suggest that a model with three strategies

[6] Given that in real life situations the same pair of players is rarely observed, I am forced to assume homogeneity of opponents.

[7] In contrast, Palacios-Huerta (2003) merges C with the natural side of the kicker for both kicker and goalkeeper.

is empirically more relevant, next I test the first implication using also action C, as Chiappori et al (2002) did. The results are shown in Table 3.4. These results show that the null hypothesis is rejected for two kickers at the 1% level and one kicker and two goalkeepers at the 5% level. Given that the probability that three or more kickers out of 9 rejecting the null at the 5% level is 0.008, I can claim that the scoring probabilities of some kickers differ depending on the action. Similarly, the goalkeepers do not behave as the theory predicts either because the probability of two or more goalkeepers out of 6 rejecting the null at the 5% is 0.033 (see again section 2.8.7).

Next, I decided to carry out a stronger test by expanding the actions of the players to LD, LU, CD, CU, RD and RU. I can only do it for kickers because goalkeepers do not seem to jump LU or RU, so that they only follow the three strategies already considered. This may happen because it is virtually impossible for a goalkeeper to jump sufficiently high when a penalty is shot in the LU or RU directions. The results, which I present in Table 3.5, show that the p-values of the LM test for kickers slightly increase compared to the ones I obtained with three actions. But qualitatively, the results obtained with six and three actions are similar, with one more kicker rejecting the null at the 10% level.

In summary, the empirical evidence on professional penalty kicks is not consistent with the implication of equal scoring probabilities for some players. In contrast, when I exclude C from the analysis, they seem to behave as the theory predicts. This means that including C seems crucial to detect possible departures from the equilibrium implications.

3.5.2 *Test for serial independence*

As I mentioned earlier, the second implication that I check is that the actions taken by the players must not be serially dependent. Palacios-Huerta (2003) uses a so-called "runs test" to evaluate this hypothesis (see Bradley (1968) for more details). A run is a sequence of consecutive identical values. If there are too many or too few runs then the serial independence hypothesis will be rejected. Too few runs means that the player does not change the action chosen often enough, which implies positive serial correlation. In contrast, too many runs means negative serial correlation.[8] However, given that runs tests do not generalize to three or more actions, I rely instead on the methods described in section 3.4.1. The results with two actions are shown in Table 3.6. This table shows the null hypotheses of serial independence with two actions is rejected for three goalkeepers at the 10% significance level, but none at the 1% or 5% levels. Hence, if

[8] The hidden Markov model in Shachat et al (2015) can also generate persistent action changes in lab games.

I rely on the usual 5% level, the evidence I obtained seems consistent with the implication of serial independence for all the kickers and goalkeepers in the sample.

As in section 3.6.1, I also expanded the actions of the players to L, C and R. The results of the numerically invariant LM tests are shown in Table 3.7.[9] In this case, the null hypothesis is not rejected for any of the players, implying that the results seem again in line with the theory. Therefore, the evidence on penalty kicks is consistent with the implication of serial independence, which is perhaps not surprising because actual penalty kicks usually take place weeks if not months apart. These findings suggest that professional soccer players seem truly able to generate random sequences; they do not appear to switch strategies too often or too seldom. This differs from the evidence of negative serial dependence in Walker and Wooders (2001), who tested whether professional tennis players played according to mixed strategies when serving and receiving.

3.6 Conclusions

In this chapter I study independence tests between two categorical variables, which only take a finite number of values H and K, respectively.

From the econometric point of view, I prove the numerical equivalence between Pearson's independence test statistic in contingency tables, the Lagrange Multiplier test statistic in several popular regression models: the multivariate LPM, the conditional and unconditional multinomial model, the multinomial logit and probit models; and the corresponding J-test statistic for overidentifying restrictions in GMM. In fact, the same results holds if one exchanges regressors and regressands in all these models. Therefore, different researchers using different econometric procedures will reach exactly the same conclusions if they use any of the aforementioned tests.

Additionally, I show that the Likelihood Ratio test statistic of independence in the conditional and unconditional multinomial model, multinomial logit and probit models are numerically identical, and that the heteroskedasticity-robust Wald test statistic in the multivariate LPM and GMM coincide with the Wald test statistic in the conditional multinomial model.

Given that the LM test statistic is numerically equivalent in all those models, all the other independence tests will also be asymptotically equivalent. Therefore, the only reason why researchers might reach different conclusions in empirical applications is because they use LR or

[9] I compute the asymptotic critical values of the LM tests using the F-approximation recommended by Stata, which is supposed to be more reliable in finite samples (see section 3.7.7 for more details).

Wald versions rather than LM tests, not because they use different models.

All these equivalences also apply to tests of serial independence of a discrete Markov chain, which can be regarded as a time series analog to the multinomial model.

From the empirical point of view, I check if professional soccer players satisfy the independence implications of mixed strategy equilibrium. To do so, I collected a dataset of 549 penalty kicks in professional soccer games that include very detailed information on many relevant aspects of the play, and specifically actions and outcomes. I find that some professional soccer players do not behave consistently with the implication of equal scoring probabilities across strategies. In contrast, I find that the second testable implication (player's actions are serially independent) holds for all of the players in the sample.

Anatolyev and Kosenok (2009) showed that Pearson goodness of fit test is also asymptotically equivalent to a multivariate regression Wald test. Additionally, Bouissou et al (1986) explain how the LR test can be used to test that a discrete Markov chain is of order k rather than $k + 1$. Extending the numerical equivalence results of this chapter to the tests considered by those authors, provides interesting avenues for subsequent research.

3.7 Proofs

3.7.1 *Proof of Proposition 3.1*

Contingency table test For my purposes, the test statistic (3.1) can be conveniently written as:

$$Pearson = n\sum_{k=1}^{K}\sum_{h=1}^{H-1}\frac{1}{n_{*k}n_{h\diamond}}\left(n_{hk}-\frac{n_{*k}n_{h\diamond}}{n}\right)^2 + n\sum_{k=1}^{K}\sum_{h=1}^{H-1}\frac{1}{n_{*k}n_{H\diamond}}\left(n_{hk}-\frac{n_{*k}n_{h\diamond}}{n}\right)^2$$
$$+2n\sum_{k=1}^{K}\sum_{h=1}^{H-1}\sum_{m=h+1}^{H-1}\left[\frac{1}{n_{*k}n_{H\diamond}}\left(n_{hk}-\frac{n_{*k}n_{h\diamond}}{n}\right)\left(n_{mk}-\frac{n_{*k}n_{m\diamond}}{n}\right)\right], \tag{3.5}$$

where $n_{h\diamond} = \sum_{k=1}^{K} n_{hk}$, $n_{*k} = \sum_{h=1}^{H} n_{hk}$, $n_{H\diamond} = n - \sum_{h=1}^{H-1} n_{h\diamond}$ and $n_{Hk} = n_{*k} - \sum_{h=1}^{H-1} n_{hk}$ for all $k = 1, \ldots, K$ and $h = 1, \ldots, H-1$.

Multivariate regression The contribution from observation i to the log-likelihood function of the multivariate regression model in (3.2) is:

$$\ln \mathcal{L}_i = -\frac{N}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}[(y_i - \Pi x_i)'\,\Sigma^{-1}\,(y_i - \Pi x_i)].$$

Following Magnus (2007), the score of the full sample can be written in matrix notation as

$$S_\delta(\Pi, \Sigma) = \left(X'Y - X'X\Pi'\right)\Sigma^{-1}. \tag{3.6}$$

Note that $Y'Y = diag(n_{1\diamond}, \cdots, n_{H-1\diamond})$, $(X'X)^{-1} = diag(n_{*1}^{-1}, \cdots, n_{*K}^{-1})$ and $X'Y$ is an $K \times (H-1)$ matrix with rows of the form $(n_{1k}, \cdots, n_{H-1,k})$, for $k = 1, \ldots, K$.

The null hypothesis of independence implies that $\delta_{h1} = \cdots = \delta_{hK} = \delta_h$. Using that $\sum_{k=1}^{K} A_{ki} = 1$, the model becomes $B_h = \delta_h + u_h$, for $h = 1, \ldots, H-1$ with $\tilde{\delta}_h = n_{h\diamond}/n$. As a result, the estimated residual covariance matrix under the null is:

$$\tilde{\Sigma}_R = \frac{1}{n}[Y'Y - Y'l_n(l_n'l_nl_n'Y] = \frac{1}{n}\begin{bmatrix} n_{1\diamond}[1-(n_{1\diamond}/n)] & \cdots & -(n_{1\diamond}n_{H-1\diamond}/n) \\ \vdots & \ddots & \vdots \\ -(n_{1\diamond}n_{H-1\diamond}/n) & \cdots & n_{H-1\diamond}[1-(n_{H-1\diamond}/n)] \end{bmatrix},$$

where l_n is an $n \times 1$ vector of ones. Note that $Y'l_n = (n_{1\diamond}, \cdots, n_{H-1\diamond})'$, $l_n'Y = (Y'l_n)'$ and $l_n'l_n = n$. Hence, $\tilde{\Sigma}_R = n^{-1}(G + wr')$, where $G = Y'Y$, $w = -Y'l_n$ and $r' = -n^{-1}w'$. Using the Sherman-Morrison (1950) formula, I get

$$\tilde{\Sigma}_R^{-1} = n\{G^{-1} - \left[\left(G^{-1}wr'G^{-1}\right)/\left(1 + r'G^{-1}w\right)\right]\},$$

so that

$$\tilde{\Sigma}_R^{-1} = n\left[\left(Y'Y\right)^{-1} + \frac{l_n l_n'}{n_{H\diamond}}\right] = \begin{bmatrix} (n/n_{1\diamond}) + (n/n_{H\diamond}) & \cdots & n/n_{H\diamond} \\ \vdots & \ddots & \vdots \\ n/n_{H\diamond} & \cdots & (n/n_{H-1\diamond}) + (n/n_{H\diamond}) \end{bmatrix}.$$

Additionally, given that under the null $\tilde{\Pi}_R' = l_K(l_n' l_n)^{-1} l_n' Y$, with l_K being a $K \times 1$ vector of ones, then

$$\left(X'Y - X'X\tilde{\Pi}_R'\right) = \begin{bmatrix} n_{11} - (n_{*1}n_{1\diamond}/n) & \cdots & n_{H-1,1} - (n_{*1}n_{H-1\diamond}/n) \\ \vdots & \ddots & \vdots \\ n_{1K} - (n_{*K}n_{1\#}/n) & \cdots & n_{H-1K} - (n_{*K}n_{H-1\diamond}/n) \end{bmatrix}.$$

Therefore, given (3.6) and the previous expressions, the element k, h of the score for $k = 1, \ldots, K$ and $h = 1, \ldots, H-1$, evaluated under the null is:

$$\begin{aligned} S_{\delta_{hk}} &= \left(n_{hk} - \frac{n_{*k}n_{h\diamond}}{n}\right)\left(\frac{n}{n_{h\diamond}} + \frac{n}{n_{H\diamond}}\right) + \sum_{z=1}^{H-1}\left[\left(n_{zk} - \frac{n_{*k}n_{z\diamond}}{n}\right)\left(\frac{n}{n_{H\diamond}}\right)\right] \\ &= n\left[(n_{hk}/n_{h\diamond}) - (n_{Hk}/n_{H\diamond})\right], \end{aligned}$$

because $\sum_{z=1}^{H-1} n_{zk} = n_{*k} - n_{hk} - n_{Hk}$ and $\sum_{z=1}^{H-1} n_{z\diamond} = n - n_{h\diamond} - n_{H\diamond}$.

Given that $vec(ABC) = (C' \otimes A)vec(B)$, then

$$vec\left[S_\delta\left(\tilde{\Pi}_R, \tilde{\Sigma}_R\right)\right] = \left(\tilde{\Sigma}_R^{-1} \otimes I\right) vec[X'Y - X'Xl_K(l_n' l_n)^{-1} l_n' Y].$$

The LM test is defined as

$$LM = vec'\left[S_\delta\left(\tilde{\Pi}_R, \tilde{\Sigma}_R\right)\right]\tilde{\mathcal{I}}_R^{-1} vec\left[S_\delta\left(\tilde{\Pi}_R, \tilde{\Sigma}_R\right)\right],$$

where $\tilde{\mathcal{I}}_R = \left[\tilde{\Sigma}_R^{-1} \otimes (X'X)\right]$, so

$$LM = vec'[X'Y - X'Xl_K(l_n' l_n)^{-1} l_n' Y]\left[\tilde{\Sigma}_R^{-1} \otimes (X'X)^{-1}\right] vec[X'Y - X'Xl_K(l_n' l_n)^{-1} l_n' Y]$$

due to the properties of the Kronecker product. Define $F = X'Y - X'Xl_K(l_n' l_n)^{-1} l_n' Y$, so

$$LM = vec'(F)\left[\tilde{\Sigma}_R^{-1} \otimes (X'X)^{-1}\right] vec(F). \tag{3.7}$$

If we expand this expression, then it immediately follows that (3.7) is the same as (3.5) for all $k = 1, \ldots, K$ and $h = 1, \ldots, H-1$.

Conditional multinomial model The derivatives of (3.3) under the alternative with respect to P_{hk} and π_{*k} are:

$$\frac{\partial \ln \mathcal{L}}{\partial P_{hk}} = \frac{n_{hk}}{P_{hk}} - \frac{n_{*k} - \sum_{h=1}^{H-1} n_{hk}}{1 - \sum_{h=1}^{H-1} P_{hk}} \quad \text{and} \quad \frac{\partial \ln \mathcal{L}}{\partial \pi_{*k}} = \frac{n_{*k}}{\pi_{*k}} - \frac{n_{*K}}{1 - \sum_{k=1}^{K-1} \pi_{*k}}$$

for $k = 1, \ldots, K$ and $h = 1, \ldots, H-1$, so the FOC yields $\hat{P}_{hk} = n_{hk}/n_{*k}$ and $\hat{\pi}_{*k} = n_{*k}/n_{*K}$.

The conditional component of the log-likelihood function under the null, which states that $H_0 : P_{hk} = P_{h\diamond}$, for $k = 1, \ldots, K$, is

$$\ln \mathcal{L} = \sum_{h=1}^{H-1} n_{h\diamond} \ln P_{h\diamond} + \left(n - \sum_{h=1}^{H-1} n_{h\diamond} \right) \ln \left(1 - \sum_{h=1}^{H-1} P_{h\diamond} \right).$$

Taking first derivatives with respect to $P_{h\diamond}$ yields:

$$\frac{\partial \ln \mathcal{L}}{\partial P_{h\diamond}} = \frac{n_{h\diamond}}{P_{h\diamond}} - \frac{n - \sum_{h=1}^{H-1} n_{h\diamond}}{1 - \sum_{h=1}^{H-1} P_{h\diamond}}.$$

Therefore, the FOC yields $\tilde{P}_{h\diamond} = n_{h\diamond}/n$, so $\tilde{P}_{h\diamond} = \tilde{\delta}_h$ for $h = 1, \ldots, H-1$.

The Hessian of this conditional multinomial model is $H(\theta) = \partial^2 \ln \mathcal{L}(\theta)/\partial\theta\theta'$, for $\theta = (P_{11}, \cdots, P_{H-1,K})$, so $H(\theta) = diag[H_1(\theta), \cdots, H_K(\theta)]$, where

$$H_k(\theta) = \begin{bmatrix} \partial^2 \ln \mathcal{L}(\theta)/\partial P_{1k}^2 & \cdots & \partial^2 \ln \mathcal{L}(\theta)/\partial P_{1k}\partial P_{H-1,k} \\ \cdots & \cdots & \cdots \\ \partial^2 \ln \mathcal{L}(\theta)/\partial P_{1k}\partial P_{H-1,k} & \cdots & \partial^2 \ln \mathcal{L}(\theta)/\partial P_{H-1,k}^2 \end{bmatrix}$$

with

$$\frac{\partial^2 \ln \mathcal{L}(\theta)}{\partial P_{hk}^2} = -\frac{n_{hk}}{P_{hk}^2} - \frac{n_{Hk}}{1 - \sum_{h=1}^{H-1} P_{hk}} \quad \text{and} \quad \frac{\partial^2 \ln \mathcal{L}(\theta)}{\partial P_{1k}\partial P_{H-1,k}} = -\frac{n_{Hk}}{\left(1 - \sum_{h=1}^{H-1} P_{hk}\right)^2},$$

for $k = 1, \ldots, K$ and $h = 1, \ldots, H-1$.

For a correctly specified likelihood, we have the information matrix equality $Var[s(\theta)] = -E[H(\theta)] = \mathcal{I}(\theta)$. The information matrix is $\mathcal{I}(\theta) = diag[\mathcal{I}_1(\theta), \cdots, \mathcal{I}_K(\theta)]$, where

$$\mathcal{I}_k(\theta) = nE(A_{ki}) \begin{bmatrix} (P_{1k})^{-1} + \left(1 - \sum_{h=1}^{H-1} P_{hk}\right)^{-1} & \cdots & \left(1 - \sum_{h=1}^{H-1} P_{hk}\right)^{-1} \\ \cdots & \cdots & \cdots \\ \left(1 - \sum_{h=1}^{H-1} P_{hk}\right)^{-1} & \cdots & (P_{H-1,k})^{-1} + \left(1 - \sum_{h=1}^{H-1} P_{hk}\right)^{-1} \end{bmatrix}.$$

Hence, using the Sherman-Morrison (1950) formula, its inverse will be given by:

$$\mathcal{I}_k(\theta)^{-1} = \frac{1}{nE(A_{ki})} \begin{bmatrix} P_{1k}(1-P_{1k}) & \cdots & -P_{1k}P_{H-1,k} \\ \cdots & \cdots & \cdots \\ -P_{H-1,k}P_{1k} & \cdots & P_{H-1,k}(1-P_{H-1,k}) \end{bmatrix}. \tag{3.8}$$

Note that the score under the null is $s(\tilde{\theta}) = n(\frac{n_{11}}{n_{1\diamond}} - \frac{n_{H1}}{n_{H\diamond}}, \cdots, \frac{n_{H-1.K}}{n_{H=1\diamond}} - \frac{n_{HK}}{n_{H\diamond}})'$, which is the same as the element h, k of the score under the null of the multivariate regression model, except that these scores are calculated by vectorizing the matrix P by columns while the ones in the multivariate regression are vectorized by rows.

Thus, one can go from one to another using the commutation matrix (see Magnus and Neudecker (1988) for more details). The most useful property of such matrix is that it allows the Kronecker products to commute. For that reason, the information matrix of the multivariate regression and the one in the multinomial model look as a mirror image of one another, i.e. $\mathcal{I} = [(X'X)\otimes\Sigma^{-1}]$ instead of $[\Sigma^{-1}\otimes(X'X)]$. Given that after performing appropriate re-ordering, the score and information matrix under the null are identical to the score and information matrix of the multivariate regression, the LM tests will also be numerically equal.

Unconditional multinomial model Ruud (2000) section 17.4 results imply that the LM test statistic is numerically invariant to non-linear transformations of the restrictions when the information matrix is used for its calculation instead of the Hessian. Consequently, the LM test for the null hypothesis $H_0 : \pi_{hk} = \pi_{*k} \times \pi_{h\diamond}$ will also be identical if one parameterizes the multinomial log-likelihood in terms of the joint probabilities π_{hk} instead of the conditional probabilities P_{hk} and the marginal ones π_{*k}. Obviously, the same is true if one uses an alternative parametrization which expresses the joint probabilities in terms of the two sets of marginal probabilities π_{*k}, for $k = 1, \dots, K-1$, and $\pi_{h\diamond}$, for $h = 1, \dots, H-1$, and $(K-1) \times (H-1)$ additional parameters ϑ which should be 0 under the null. In particular, in the 2×2 case, the log-likelihood written in this way is just

$$\ln \mathcal{L} = n_{11} \ln[(1-\pi_{*2})(1-\pi_{2\diamond}) + \vartheta] + n_{12} \ln[(1-\pi_{2\diamond})\pi_{*2} - \vartheta]$$
$$+n_{21} \ln[(1-\pi_{*2})\pi_{2\diamond} - \vartheta] + n_{22} \ln(\pi_{*2}\pi_{2\diamond} + \vartheta).$$

The score vector for $\theta = (\pi_{*2}, \pi_{2\diamond}, \vartheta)$ is defined as

$$\begin{aligned}
\frac{\partial \ln \mathcal{L}}{\partial \pi_{*2}} &= -\frac{n_{11}\left(1-\pi_{2\diamond}\right)}{\left(1-\pi_{*2}\right)\left(1-\pi_{2\diamond}\right)+\vartheta} + \frac{n_{12}\left(1-\pi_{2\diamond}\right)}{\left(1-\pi_{2\diamond}\right)\pi_{*2}-\vartheta} - \frac{n_{21}\pi_{2\diamond}}{\left(1-\pi_{*2}\right)\pi_{2\diamond}-\vartheta} + \frac{n_{22}\pi_{2\diamond}}{\pi_{*2}\pi_{2\diamond}+\vartheta}\\
\frac{\partial \ln \mathcal{L}}{\partial \pi_{2\diamond}} &= -\frac{n_{11}\left(1-\pi_{*2}\right)}{\left(1-\pi_{*2}\right)\left(1-\pi_{2\diamond}\right)+\vartheta} - \frac{n_{12}\pi_{*2}}{\left(1-\pi_{2\diamond}\right)\pi_{*2}-\vartheta} + \frac{n_{21}\left(1-\pi_{*2}\right)}{\left(1-\pi_{*2}\right)\pi_{2\diamond}-\vartheta} + \frac{n_{22}\pi_{*2}}{\pi_{*2}\pi_{2\diamond}+\vartheta}\\
\frac{\partial \ln \mathcal{L}}{\partial \vartheta} &= \frac{n_{11}}{\left(1-\pi_{*2}\right)\left(1-\pi_{2\diamond}\right)+\vartheta} - \frac{n_{12}}{\left(1-\pi_{2\diamond}\right)\pi_{*2}-\vartheta} - \frac{n_{21}}{\left(1-\pi_{*2}\right)\pi_{2\diamond}-\vartheta} + \frac{n_{22}}{\pi_{*2}\pi_{2\diamond}+\vartheta}
\end{aligned}$$

so the FOC yields $\hat{\pi}_{*2} = n_{*2}/n$, $\hat{\pi}_{2\diamond} = n_{2\diamond}/n$ and $\hat{\vartheta} = n^{-1}n_{22} - (\hat{\pi}_{*2}\hat{\pi}_{2\diamond})$, where $n_{*2} = n_{12}+n_{22}$, $n_{2\diamond} = n_{21} + n_{22}$ and $n = n_{11} + n_{12} + n_{21} + n_{22}$. Therefore, the unrestricted estimators of the marginal probabilities ($\hat{\pi}_{*2}$ and $\hat{\pi}_{2\diamond}$) coincide with the restricted ones ($\tilde{\pi}_{*2}$ and $\tilde{\pi}_{2\diamond}$).

Moreover, the Hessian of the log-likelihood function for $\theta = (\pi_{*2}, \pi_{2\diamond}, \vartheta)$ is

$$\begin{aligned}
\frac{\partial^2 \ln \mathcal{L}}{\partial \pi_{*2}^2} &= -\frac{n_{11}\left(1-\pi_{2\diamond}\right)^2}{\left[\left(1-\pi_{*2}\right)\left(1-\pi_{2\diamond}\right)+\vartheta\right]^2} - \frac{n_{12}\left(1-\pi_{2\diamond}\right)^2}{\left[\left(1-\pi_{2\diamond}\right)\pi_{*2}-\vartheta\right]^2} - \frac{n_{21}\pi_{2\diamond}^2}{\left[\left(1-\pi_{*2}\right)\pi_{2\diamond}-\vartheta\right]^2} - \frac{n_{22}\pi_{2\diamond}^2}{\left(\pi_{*2}\pi_{2\diamond}+\vartheta\right)^2}\\
\frac{\partial^2 \ln \mathcal{L}}{\partial \pi_{*2}\pi_{2\diamond}} &= \frac{n_{11}\vartheta}{\left[\left(1-\pi_{*2}\right)\left(1-\pi_{2\diamond}\right)+\vartheta\right]^2} + \frac{n_{12}\vartheta}{\left[\left(1-\pi_{2\diamond}\right)\pi_{*2}-\vartheta\right]^2} + \frac{n_{21}\vartheta}{\left[\left(1-\pi_{*2}\right)\pi_{2\diamond}-\vartheta\right]^2} + \frac{n_{22}\vartheta}{\left(\pi_{*2}\pi_{2\diamond}+\vartheta\right)^2}\\
\frac{\partial^2 \ln \mathcal{L}}{\partial \pi_{*2}\vartheta} &= \frac{n_{11}\left(1-\pi_{2\diamond}\right)}{\left[\left(1-\pi_{*2}\right)\left(1-\pi_{2\diamond}\right)+\vartheta\right]^2} + \frac{n_{12}\left(1-\pi_{2\diamond}\right)}{\left[\left(1-\pi_{2\diamond}\right)\pi_{*2}-\vartheta\right]^2} - \frac{n_{21}\pi_{2\diamond}}{\left[\left(1-\pi_{*2}\right)\pi_{2\diamond}-\vartheta\right]^2} - \frac{n_{22}\pi_{2\diamond}}{\left(\pi_{*2}\pi_{2\diamond}+\vartheta\right)^2}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ln \mathcal{L}}{\partial \pi_{2\diamond}^2} &= -\frac{n_{11}\left(1-\pi_{*2}\right)^2}{\left[\left(1-\pi_{*2}\right)\left(1-\pi_{2\diamond}\right)+\vartheta\right]^2} - \frac{n_{12}\pi_{*2}^2}{\left[\left(1-\pi_{2\diamond}\right)\pi_{*2}-\vartheta\right]^2} - \frac{n_{21}\left(1-\pi_{*2}\right)^2}{\left[\left(1-\pi_{*2}\right)\pi_{2\diamond}-\vartheta\right]^2} - \frac{n_{22}\pi_{*2}^2}{\left(\pi_{*2}\pi_{2\diamond}+\vartheta\right)^2}\\
\frac{\partial^2 \ln \mathcal{L}}{\partial \pi_{2\diamond}\vartheta} &= \frac{n_{11}\left(1-\pi_{*2}\right)}{\left[\left(1-\pi_{*2}\right)\left(1-\pi_{2\diamond}\right)+\vartheta\right]^2} - \frac{n_{12}\pi_{*2}}{\left[\left(1-\pi_{2\diamond}\right)\pi_{*2}-\vartheta\right]^2} + \frac{n_{21}\left(1-\pi_{*2}\right)}{\left[\left(1-\pi_{*2}\right)\pi_{2\diamond}-\vartheta\right]^2} - \frac{n_{22}\pi_{*2}}{\left(\pi_{*2}\pi_{2\diamond}+\vartheta\right)^2}\\
\frac{\partial^2 \ln \mathcal{L}}{\partial \vartheta^2} &= -\frac{n_{11}}{\left[\left(1-\pi_{*2}\right)\left(1-\pi_{2\diamond}\right)+\vartheta\right]^2} - \frac{n_{12}}{\left[\left(1-\pi_{2\diamond}\right)\pi_{*2}-\vartheta\right]^2} - \frac{n_{21}}{\left[\left(1-\pi_{*2}\right)\pi_{2\diamond}-\vartheta\right]^2} - \frac{n_{22}}{\left(\pi_{*2}\pi_{2\diamond}+\vartheta\right)^2}.
\end{aligned}$$

Under the null, which states that $\vartheta = 0$, the information matrix is

$$I\left(\theta\right) = n \times diag\left\{ \begin{array}{ccc} \left[\left(1-\pi_{*2}\right)\pi_{*2}\right]^{-1} & \left[\left(1-\pi_{2\diamond}\right)\pi_{2\diamond}\right]^{-1} & \left[\left(1-\pi_{*2}\right)\left(1-\pi_{2\diamond}\right)\pi_{*2}\pi_{2\diamond}\right]^{-1} \end{array} \right\}. \quad (3.9)$$

The LM test statistic is defined by $LM = n \times s(\tilde{\theta})' I(\tilde{\theta})^{-1} s(\tilde{\theta})$, where $s(\tilde{\theta})$ and $I(\tilde{\theta})$ are the unrestricted gradient and information matrix evaluated at the restricted estimators $\tilde{\theta}$, respectively. Given (3.9), the estimated information matrix $I(\tilde{\theta})$ will be

$$I(\tilde{\theta}) = n \times diag\left[\begin{array}{ccc} n^2\left(n_{*1}n_{*2}\right)^{-1} & n^2\left(n_{1\diamond}n_{2\diamond}\right)^{-1} & n^4\left(n_{*1}n_{*2}n_{1\diamond}n_{2\diamond}\right)^{-1} \end{array} \right],$$

where $n_{*1} = 1 - n_{*2}$ and $n_{1\diamond} = 1 - n_{2\diamond}$. Additionally, the score under the null is given by

$$s(\tilde{\theta}) = n^{-1}\left(\begin{array}{ccc} 0 & 0 & s_{11}/s_{12} \end{array} \right),$$

where

$$s_{11} = n^2 \left(n_{11}^2 n_{22} - n_{11} n_{12} n_{21} + n_{11} n_{21} n_{22} + n_{11} n_{22}^2 - n_{12}^2 n_{21} - n_{12} n_{21}^2 - n_{12} n_{21} n_{22} \right)$$

and

$$s_{12} = (n_{11} + n_{12})(n_{11} + n_{21})(n_{12} + n_{22})(n_{21} + n_{22}).$$

Therefore, the LM test statistic will be

$$LM = n \times \frac{\hat{\vartheta}^2}{(1 - \hat{\pi}_{*2})(1 - \hat{\pi}_{2\diamond})\, \hat{\pi}_{*2} \hat{\pi}_{2\diamond}},$$

where $\hat{\vartheta}$ is defined above. Hence, the LM test, which is exactly the same as (3.5), is like a $t-$test for ϑ, but using the asymptotic standard error computed under the null.

The same procedure can be applied to the general $H \times K$ case. For example, in a 3×3 contingency table, the joint probabilities can be parameterized in terms of the marginal probabilities $\pi_{h\diamond}$ and π_{*k}, for $h, k = 1, 2, 3$, and four additional parameters ϑ_{11}, ϑ_{12}, ϑ_{21} and ϑ_{22}, which should be 0 under the null.

Analogous derivations show that the estimators of the marginal probabilities are the same under the null and the alternative, and that the information matrix evaluated under the null is block diagonal between $\pi_{h\diamond}$, π_{*k} and the $\vartheta' s$.

Multinomial logit model Recall that the log-likelihood function of this model is:

$$\ln \mathcal{L}(\gamma) = \sum_{k=1}^{K} \left[\sum_{h=1}^{H-1} n_{hk} \ln P_{hk} + n_{Hk} \ln \left(1 - \sum_{h=1}^{H-1} P_{hk} \right) \right].$$

The score is defined as $s(\gamma) = \partial \mathcal{L}(\gamma)/\partial \gamma$ with

$$\frac{\partial \ln \mathcal{L}(\gamma)}{\partial P_{hk}} = \left[\frac{n_{hk}}{P_{hk}} - \frac{n_{Hk}}{1 - \sum_{h=1}^{H-1} P_{hk}} \right] \quad \text{and} \quad \frac{\partial P_{hk}}{\partial \gamma_{hk}} = \frac{\exp(\gamma_{hk} A_{ki}) A_{ki}}{\left[1 + \exp(\gamma_{hk} A_{ki}) + \exp(\gamma_{kh} A_{ki})\right]^2},$$

for $k = 1, \ldots, K$ and $h = 1, \ldots, H-1$. Solving the FOC for $\hat{P}_{hk}$ yields $\hat{P}_{hk} = n_{hk}/n_{*k}$, which is again the same as the estimate in the multivariate regression, with $\hat{\gamma}_{hk} = \ln(n_{hk}) - \ln(n_{*k} - n_{hk})$ (see Cameron and Trivedi (2005) chapter 15, section 4 for more details).

The multinomial logit log-likelihood function under the null hypothesis $H_0 : P_{hk} = P_h$, for $k = 1, \ldots, K$ and $h = 1, \ldots, H-1$, is:

$$\ln \mathcal{L}(\gamma) = \sum_{k=1}^{K} \left\{ \sum_{h=1}^{H-1} \left[(n_{hk} \ln P_h) + n_{Hk} \ln \left(1 - \sum_{h=1}^{H-1} P_h \right) \right] + n_{*k} \ln \pi_k \right\},$$

which yields $\tilde{P}_h = n_{h\diamond}/n$, given that $\sum_{h=1}^{K} n_{h\diamond} = n$. Additionally, $\tilde{P}_h = \exp(\tilde{\gamma}_h)/\sum \exp(\tilde{\gamma}_h)$.

It is worth mentioning that the values of the log-likelihood function under the null and alternative of the multinomial logit are equal to the corresponding log-likelihoods of the multinomial model, which implies that the LM and LR test statistics will be the same (see section 17.4 of Ruud (2000)).

Multinomial probit model Recall that the observation rule is

$$B_{hi} = 1\left\{B_{hi}^* = \max_{j=1,\ldots,n} B_{hj}^*\right\},$$

where $1\{\}$ is the indicator function, so that $B_{hi} = 1$ if the h is the preferred choice. Otherwise, B_{hi} equals zero. Therefore, the log-likelihood function is

$$L(\theta) = \sum_{i=1}^{n}\sum_{h=1}^{H} B_{hi} \ln \Pr(B_{hi} = 1|x).$$

Like in the multinomial logit, this log-likelihood function coincides with a complicated nonlinear reparametrization of the conditional component of the log-likelihood function of the multinomial model.

For simplicity of exposition, consider the $2 \times K$ case. The log-likelihood function is:

$$\ln \mathcal{L} = \sum_{i=1}^{n}\{B_i \ln \Phi\left(\alpha_1 A_{1i} + \cdots + \alpha_K A_{Ki}\right) + (1 - B_i)\ln[1 - \Phi\left(\alpha_1 A_{1i} + \cdots + \alpha_K A_{Ki}\right)]\},$$

where $\Phi(\cdot)$ is the standard normal cdf. The score is defined as $s(\alpha) = \partial \ln \mathcal{L}(\alpha)/\partial\alpha$ with

$$\frac{\partial \ln \mathcal{L}(\alpha)}{\partial \alpha_k} = \frac{A_k[\phi\left(\alpha_k A_k\right)](B_i - P_k)}{P_k(1 - P_k)},$$

where $P_k = \Phi\left(\alpha_1 A_{1i} + \cdots + \alpha_K A_{Ki}\right)$, for $k = 1, \ldots, K$. Solving for $\hat{P}_k$ yields $\hat{P}_k = n_k/n_{*k}$ which is again the same as in the conditional multinomial model, with $\hat{\alpha}_k = \Phi^{-1}\left(n_k/n_{*k}\right)$.

In turn, the log-likelihood function under $H_0 : \alpha_k = \alpha$, for $k = 1, \ldots, K$, is:

$$\ln \mathcal{L} = \sum_{i=1}^{n}\{B_i \ln \Pr(B_i = 1) + (1 - B_i)\ln[1 - \Pr(B_i = 1)]\},$$

which yields $\widehat{\Pr}(B = 1) = n_{1\diamond}/n$. As a result, the multinomial probit model under the null is entirely analogous to the multinomial logit one, so the same numerical equalities hold.

GMM As explained in section 3.2, the moment conditions are (3.4). Under H_1, Π is unrestricted while under H_0, $\Pi(\upsilon) = \upsilon l_K'$. The GMM estimator of υ is defined as:

$$\tilde{\upsilon} = \arg\min_{\upsilon} \left(\frac{1}{n} \sum_{i=1}^{n} \{[y_i - \Pi(\upsilon)x_i] \otimes x_i\} \right)' \Upsilon^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \{[y_i - \Pi(\upsilon)x_i] \otimes x_i\} \right),$$

where Υ is a symmetric positive definite $[K \times (H-1)] \times [K \times (H-1)]$ weight matrix.

The GMM estimator can also be written as $\tilde{\upsilon} = \arg\min_{\upsilon} \bar{g}'\Upsilon^{-1}\bar{g}$, where $\bar{g} = \frac{1}{n}\sum_{i=1}^{n}\{g[z_i; \Pi(\upsilon)]\}$. Hence, the FOC is $2\bar{g}'\Upsilon^{-1}\left(\partial\bar{g}/\partial\hat{\upsilon}'\right) = 0$, with $\Upsilon = [\Sigma_R \otimes \sum_{i=1}^{n}(x_i x_i')]$ being optimal under H_0.

Given that (3.4) are linear, one can rewrite $\bar{g}$ as $\bar{g} = \bar{m}_n - \bar{M}_n\upsilon$, with $\bar{M}_n = \frac{1}{n}\sum_{i=1}^{n}(I_{H-1} \otimes x_i)$ and $\bar{m}_n = \frac{1}{n}\sum_{i=1}^{n}(y_i \otimes x_i)$, which implies that $\tilde{\upsilon} = \left(\bar{M}_n\Upsilon^{-1}\bar{M}_n\right)'\left(\bar{M}_n\Upsilon^{-1}\bar{m}_n\right)$. Specifically,

$$\tilde{\upsilon} = \left\{ \left[\frac{1}{n}\sum_{i=1}^{n}(I_{H-1} \otimes x_i)\right]' \left[\tilde{\Sigma}_R \otimes \frac{1}{n}\sum_{i=1}^{n}(x_i x_i')\right]^{-1} \left[\frac{1}{n}\sum_{i=1}^{n}(I_{H-1} \otimes x_i)\right] \right\}^{-1} = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

whose representative element $n_{h\diamond}/n$ is exactly the same as the restricted estimator $\tilde{\delta}_h$ in the multivariate regression.

The J-test statistic for overidentifying restrictions is $J = n \times \bar{g}(\tilde{\upsilon})'\Upsilon^{-1}\bar{g}(\tilde{\upsilon})$. Note that $\Upsilon^{-1} = \Sigma_R^{-1} \otimes (X'X)^{-1}$ is exactly the same as the information matrix in the multivariate regression. Additionally,

$$g_i = (y_i - \Pi x_i) \otimes x_i = (y_i \otimes x_i)\, l_K - (\Pi x_i \otimes x_i)\, l_K = vec(x_i l_K y_i') - vec(x_i l_k x_i'\Pi')$$

and since $\Pi_R(\upsilon) = \upsilon l_K' = (l_K \upsilon' I_{H-1})'$, then $g_i = vec(x_i y_i') - (I_{H-1} \otimes x_i x_i')\,\upsilon$ with $\delta = (I_{H-1} \otimes l_k)\, vec(\upsilon) = (I_{H-1} \otimes l_K)\,\upsilon$, so $g_i = vec(x_i y_i') - (I_{H-1} \otimes l_k)\,\upsilon$. This implies that

$$\frac{1}{n}\sum_{i=1}^{n} g_i = \bar{g}(z; \tilde{\upsilon}) = vec[X'Y - X'X l_K (l_n' l_n)^{-1} l_n Y],$$

which is exactly the same as the normal equations of the multivariate regression (3.7). Therefore, the J-test statistic for overidentifying restrictions is also numerically equivalent to the LM test.

Finally, given that the model under the alternative is exactly identified, the Distance Difference test (see Newey and West (1987) for more details) is exactly the same as the J-test statistic. Hence, following the results in chapter 22 of Ruud (2000), the minimum chi-square test that compares $\hat{\Pi}$ with $\Pi(\hat{\upsilon})$ and the GMM version of the LM test will also be numerically identical to the J-test statistic.

3.7.2 *Proof of Proposition 3.2*

The results in section 17.4 of Ruud (2000) imply that the LR test of the conditional and unconditional multinomial model, the multinomial probit model and the logit model in section 3.2 must coincide because LR tests are numerically invariant to non-linear transformations of parameters and restrictions.

3.7.3 *Proof of Proposition 3.3*

Multivariate LPM The covariance matrix of the heteroskedasticity-robust Wald test in the multivariate regression is defined as $Q = (I_{H-1} \otimes X'X)^{-1}\hat{\Psi}(I_{H-1} \otimes X'X)^{-1}$, where $\hat{\Psi} = \sum_i [(u_i \otimes x_i)(u_i' \otimes x_i')]$. Note that both $\hat{\Psi}$ and $(I_{H-1} \otimes X'X)^{-1} = I_{H-1} \otimes (X'X)^{-1}$ are $[(H-1) \times K] \times [(H-1) \times K]$ matrices.

Specifically, $\hat{\Psi} = \sum_i [(u_i u_i') \otimes (x_i x_i')]$, with $x_i x_i' = \begin{pmatrix} A_{1i}^2 & \cdots & A_{1i}A_{Ki} \\ \vdots & \ddots & \vdots \\ A_{1i}A_{Ki} & \cdots & A_{Ki}^2 \end{pmatrix}$, where A_{ki}, for $k = 1, \ldots, K$, are the mutually exclusive dummy variables defined in section 2 and the u's are the regression residuals for equation i. Hence, when $A_{ki} = 1$,

$$\begin{aligned} u_{hi}^2 &= \left(B_{hi} - \tfrac{n_{hk}}{n_{*k}}\right)^2 = B_{hi}\left(1 - \tfrac{2n_{hk}}{n_{*k}}\right) + \left(\tfrac{n_{hk}}{n_{*k}}\right)^2, \\ u_{hi}u_{mi} &= \left(B_{hi} - \tfrac{n_{hk}}{n_{*k}}\right)\left(B_{mi} - \tfrac{n_{mk}}{n_{*k}}\right) = \tfrac{n_{hk}n_{mk}}{n_{*k}^2} - B_{hi}\tfrac{n_{mk}}{n_{*k}} - B_{mi}\tfrac{n_{hk}}{n_{*k}} \end{aligned}$$

and $x_i x_i'$ is a matrix of zeros except for a 1 in the i^{th} diagonal element because B_{hi} and B_{mi} are also dummy variables for $h, m = 1, \ldots, H-1$ and $h \neq m$. Therefore,

$$
\hat{\Psi} = \sum_{A_1=1} \left[\begin{matrix}
B_{hi}\left(1 - \frac{2n_{h1}}{n_{*1}}\right) + \left(\frac{n_{h1}}{n_{*1}}\right)^2 & 0 & ... & \frac{n_{h1}n_{m1}}{n_{*1}^2} - B_{hi}\frac{n_{m1}}{n_{*1}} - B_{mi}\frac{n_{h1}}{n_{*1}} & 0 \\
0 & 0 & ... & 0 & 0 \\
... & ... & ... & ... & ... \\
0 & ... & ... & ... & ... \\
\frac{n_{h1}n_{m1}}{n_{*1}^2} - B_{hi}\frac{n_{m1}}{n_{*1}} - B_{mi}\frac{n_{h1}}{n_{*1}} & 0 & ... & B_{mi}\left(1 - \frac{2n_{m1}}{n_{*1}}\right) + \left(\frac{n_{m1}}{n_{*1}}\right)^2 & 0 \\
0 & 0 & ... & 0 & 0 \\
... & ... & ... & ... & ...
\end{matrix} \right] + ...
$$

$$
+ \sum_{A_K=1} \left[\begin{matrix}
0 & ... & ... & ... & ... & ... \\
... & ... & 0 & ... & ... & 0 \\
... & 0 & B_{hi}\left(1 - \frac{2n_{hK}}{n_{*K}}\right) + \left(\frac{n_{hK}}{n_{*K}}\right)^2 & 0 & 0 & \frac{n_{hK}n_{mK}}{n_{*K}^2} - B_{hi}\frac{n_{mK}}{n_{*K}} - B_{mi}\frac{n_{hK}}{n_{*K}} \\
... & ... & 0 & ... & ... & 0 \\
... & 0 & ... & ... & ... & ... \\
... & 0 & \frac{n_{hK}n_{mK}}{n_{*K}^2} - B_{hi}\frac{n_{mK}}{n_{*K}} - B_{mi}\frac{n_{hK}}{n_{*K}} & ... & ... & B_{mi}\left(1 - \frac{2n_{mK}}{n_{*K}}\right) + \left(\frac{n_{mK}}{n_{*K}}\right)^2
\end{matrix} \right],
$$

which can be simplified to

$$\hat{\Psi} = \begin{bmatrix} n_{11}\left(1-\frac{n_{11}}{n_{*1}}\right) & 0 & ... & ... & ... & -\frac{n_{H-2,1}n_{H-1,1}}{n_{*1}} & 0 & ... \\ 0 & ... & 0 & ... & ... & 0 & ... & 0 \\ ... & 0 & n_{1K}\left(1-\frac{n_{1K}}{n_{*K}}\right) & 0 & ... & ... & 0 & -\frac{n_{H-2,K}n_{H-1,K}}{n_{*K}} \\ 0 & ... & ... & ... & ... & ... & ... & ... \\ -\frac{n_{11}n_{H-1,1}}{n_{*1}} & 0 & ... & ... & ... & n_{H-1,1}\left(1-\frac{n_{H-1,1}}{n_{*1}}\right) & 0 & ... \\ 0 & ... & 0 & ... & ... & 0 & ... & 0 \\ ... & 0 & -\frac{n_{1K}n_{H-1,K}}{n_{*K}} & ... & ... & ... & 0 & n_{H-1,K}\left(1-\frac{n_{H-1,K}}{n_{*K}}\right) \end{bmatrix}$$

because $\sum_{A_{ki}=1} B_{hk} = n_{hk}$ and $\sum_{A_{ki}=1} 1 = n_{*k}$ for $k = 1, \ldots, K$.

Therefore,

$$Q = \begin{bmatrix} \frac{n_{11}}{n_{*1}^2}\left(1-\frac{n_{11}}{n_{*1}}\right) & 0 & \dots & \dots & -\frac{n_{11}n_{H-1,1}}{n_{*1}^3} & 0 & \dots \\ 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ \dots & 0 & \frac{n_{1K}}{n_{*K}^2}\left(1-\frac{n_{1K}}{n_{*K}}\right) & \dots & \dots & 0 & -\frac{n_{1K}n_{H-1,K}}{n_{*K}^3} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ -\frac{n_{11}n_{H-1,1}}{n_{*1}^3} & 0 & \dots & \dots & \frac{n_{H-1,1}}{n_{*1}^2}\left(1-\frac{n_{H-1,1}}{n_{*1}}\right) & 0 & \dots \\ 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ \dots & 0 & -\frac{n_{1K}n_{H-1,K}}{n_{*K}^3} & \dots & \dots & 0 & \frac{n_{H-1,K}}{n_{*K}^2}\left(1-\frac{n_{H-1,K}}{n_{*K}}\right) \end{bmatrix} \quad (3.10)$$

which has $(H-1) \times (H-1)$ blocks of size K, each of which is diagonal.

Conditional multinomial model The inverse of the estimated information matrix is given by (3.8). In practice, we use the unrestricted ML estimator of P_{hk}, for $k = 1, \dots, K$ and $h = 1, \dots, H-1$ to estimate the inverse information matrix. Given that $\hat{P}_{hk} = n_{hk}/n_{*k}$ and $\hat{E}(A_{ki}) = n_{*k}/n$, then

$$\mathcal{I}_k(\hat{\theta})^{-1} = \frac{1}{n\hat{E}(A_{ki})} \begin{bmatrix} \frac{n_{1k}}{n_{*k}}(1-\frac{n_{1k}}{n_{*k}}) & \cdots & -\frac{n_{1k}}{n_{*k}}\frac{n_{H-1,k}}{n_{*k}} \\ \vdots & \ddots & \vdots \\ -\frac{n_{1k}}{n_{*k}}\frac{n_{H-1,k}}{n_{*k}} & \cdots & \frac{n_{H-1,k}}{n_{*k}}(1-\frac{n_{H-1,k}}{n_{*k}}) \end{bmatrix},$$

which is exactly the same as the heteroskedasticity-robust variance in the multivariate regression (3.10), except for re-ordering. Therefore, given that both the point estimators and the covariance matrices are the same, the Wald tests will also be the same.

GMM As I explained in section 3.2, the moment conditions are given by (3.4). Under H_1, Π is unrestricted, so the GMM estimator is defined as:

$$\hat{\Pi} = \arg\min_{\Pi} \left\{ \frac{1}{n} \sum_{i=1}^{n} [(y_i - \Pi x_i) \otimes x_i] \right\}' \Upsilon^{-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} [(y_i - \Pi x_i) \otimes x_i] \right\},$$

where Υ is a symmetric positive definite $[K \times (H-1)] \times [K \times (H-1)]$ weight matrix. However, given that the model is exactly identified under the alternative, Υ becomes irrelevant. Therefore, the GMM unrestricted estimator will be the one that sets

$$\frac{1}{n} \sum_{i=1}^{n} [(y_i - \Pi x_i) \otimes x_i] = 0.$$

But this expression is exactly the same as the score (3.6) in the multivariate LPM, which implies that the GMM unrestricted estimator coincides with OLS.

Finally, the heteroskedasticity-robust estimated variance-covariance matrix will also be exactly the same as in OLS (see section 21.4.3 of Ruud (2000)). Therefore, the Wald test statistic will coincide.

3.7.4 *Finite sample distribution*

Even in experimental studies, few observations for each player are likely to be the rule rather than the exception. Therefore, the asymptotic $\mathcal{X}^2$ distribution of the independence tests described in section 3.2 may be unreliable when the number of observations is small.

Given the discrete nature of the random variables involved, permutation-type tests may seem to provide a way of conducting exact inference to test for independence. In fact, Fisher's exact test (see Fisher (1922)) could be regarded as a permutation test. His test takes the values of $n_{*1}, \ldots, n_{*K}$ and $n_{1\diamond}, \ldots, n_{H\diamond}$ as given and therefore, it is equivalent to using the likelihood in the unconditional multinomial model in section 3.2 but treating the estimated values of $\pi_{h\diamond}$ and π_{*k} as if they were the true values of the parameters. However, this test is only exact if the marginal probabilities of the two categorical random variables are known, as in Fisher's famous tea cup classification example. In more realistic situations, those marginal probabilities are unknown, and the supposedly exact test is only valid asymptotically.

Although Monte Carlo simulations can help us in assessing how good the asymptotic approximation of a test statistic is, in practice, they are not useful for inferences in a given sample because we do not know the true values of the parameters.

For that reason, simulation methods provide an alternative to asymptotic approximations for obtaining p-values by resampling methods. Next, I explain how to compute the parametric bootstrap distribution for the penalty kick case.

Note that tests of the first and second implications of mixed strategy equilibrium with two actions are equivalent in this context because, although the variables involved are different, they are all based on 2×2 contingency tables. For that reason, I focus on the first hypothesis only.

Given that all the variables used are discrete, the number of states of the world is finite (2 possible actions per player $\times$ 2 possible outcomes per player's actions). In addition, the number of values of the estimators and test statistics will be repeated in many of those states of the world. For example, for $n = 5$ and two actions per player, there are $(2^3)^5$ possible states, but only 56 different contingency tables, while for $n = 20$ there are $(2^3)^{20}$ states, but only 1771 contingency tables (the number of possible contingency tables for $n = 5$ and $n = 20$ is obtained after carefully considering how many different values can n_{SL} take for each combination of n_L and n_S). For that reason, I simulate the contingency tables directly, which contain all the information.

Recall that in the contingency table in section 3.2.1 applied to the penalty kick case, where now $\tilde{y}$ is the outcome (success (S) or failure (F)) and x is the action of the player (L and R), both $n_L = n_{SL} + n_{FL}$ and $n_S = n_{SL} + n_{SR}$ have values that go from 0 to n. Given those values, I only need to choose an additional element to complete the contingency table. Without loss of generality, I choose n_{SL}. For fixed n_L and n_S, n_{SL} fluctuates between a maximum and a minimum. It is easy to see that the minimum value n_{SL} can take is the maximum of 0 and $n_L + n_S - n$, while the maximum value it takes is the minimum of n_L and n_S.

To find the exact probability of each of those contingency tables and therefore of the corresponding test statistics, first note that under the null hypothesis the number of kicks to the left (n_L) and the number of goals scored (n_S) are independent random variables. Therefore,

$$\Pr(n_L, n_S, n_{SL}|n) = \Pr(n_S|n) \times \Pr(n_L|n) \times \Pr(n_{SL} \mid n_L, n_S; n),$$

where $\Pr(n_j|n)$, for $j = S, L$, is binomial, whose values depend on the values of n and $E(L) = \pi_L$ or $E(S) = \pi_S$. In turn, Fisher (1922) showed that $\Pr(n_{SL} \mid n_L, n_S; n)$ is hypergeometric, with values that only depend on n, n_L and n_S. The binomial distribution gives the probability of k successes in n trials with replacement, while the hypergeometric distribution does the same thing, but without replacement. Interestingly, the probability of those contingency tables under the null is identical to the likelihood written in terms of $\pi_{hk} = P_{hk} \times \pi_{*k}$ as stated in formula (38) of Mood et al (1974).

For a given sample, I calculate n_{SL} and n_S to estimate the marginal probabilities $\hat{\pi}_{*L}$ and $\hat{\pi}_{S\diamond}$. Then I use those estimated values to independently draw the actions of the kicker, as well as whether or not the goal is scored.

However, I have found that many of those distributions are repeated for different values of $\hat{\pi}_{*L}$ and $\hat{\pi}_{S\diamond}$. More precisely, for an even number of observations there are

$$\frac{1}{2}\left(\frac{n+2}{2} \times \frac{n}{2}\right) + 1$$

different distributions, while for an odd number of observations there are

$$\frac{1}{2}\left(\frac{n-1}{2} \times \frac{n+1}{2}\right) + 1.$$

This is due to the symmetry relationships that arise because the variables of the model are mutually exclusive dummy variables. For example, $n_L = 2$ and $n_S = 3$ will give the same distribution for the test statistics as $n_L = 3$ and $n_S = 2$. As a result, there are only 7 and 56 possible different distributions for Pearson's test for $n = 5$ and $n = 20$ respectively, even though there are 56 and 1771 contingency tables.

Therefore, the only difference between the Fisher test and the simulated test distribution that I compute is that the former uses the unknown probabilities while the latter uses the estimated probabilities.

In the case of three or more actions, I can find the exact probability of each possible contingency table, and therefore the exact probability of the different test statistics, using

$$\begin{aligned}\Pr(\text{contingency table}|n) = \Pr(n_{*1}, \ldots, n_{*K}|n) \times \Pr(n_{1\diamond}, \ldots, n_{H\diamond}|n) \\ \times \Pr(n_{hk}|n_{*1}, \ldots, n_{*K}; n_{1\diamond}, \ldots, n_{H\diamond}; n),\end{aligned}$$

because $n_{h\diamond}$ and n_{*k} are independent under the null.

Given that

$$\Pr(n_{hk}|n_{*1},\ldots,n_{*K};n_{1\diamond},\ldots,n_{H\diamond};n) = \frac{(n_{*1}!,\ldots,n_{*K}!)\,(n_{1\diamond}!,\ldots,n_{H\diamond}!)}{n!\prod_{h=1}^{H}\prod_{k=1}^{K} n_{ij}!}$$
$$\Pr(n_{*1},\ldots,n_{*K}|n) = \frac{n!}{n_{*1}!,\ldots,n_{*K}!}\prod_{k=1}^{K}\pi_{*k}^{n_{*k}}$$
$$\Pr(n_{1\diamond}!,\ldots,n_{H\diamond}!|n) = \frac{n!}{n_{1\diamond}!,\ldots,n_{H\diamond}!}\prod_{h=1}^{H}\pi_{h\diamond}^{n_{h\diamond}},$$

then

$$\Pr(\text{contingency table}|n) = \frac{n!}{\prod_{h=1}^{H}\prod_{k=1}^{K} n_{ij}!}\left(\prod_{k=1}^{K}\pi_{*k}^{n_{*k}}\right)\left(\prod_{h=1}^{H}\pi_{h\diamond}^{n_{h\diamond}}\right).$$

As in the 2×2 case, the probability of those contingency tables under the null is identical to the likelihood written in terms of $\pi_{hk} = P_{hk}\times\pi_{*k}$ as stated in formula (38) of Mood et al (1974).

However, the number of possible contingency tables is very large, and finding their exact bootstrap distribution is very tedious. For that reason, I compute the p-value using Monte Carlo simulations rather than the exact test, using once again the estimated values of the marginal probabilities.

3.7.5 *Relationship Between Test Statistics When* $H = 2$ *and* $H = 3$

Following Stewart (1995), the Wald, LR and LM tests in the multivariate LPM can be written as functions of the eigenvalues $(\lambda_1,\ldots,\lambda_{H-1})$ of the matrix GE^{-1}, where $G = \tilde{\Sigma}_R'\tilde{\Sigma}_R - \hat{\Sigma}_U'\hat{\Sigma}_U$ and $E = \hat{\Sigma}_U'\hat{\Sigma}_U$, with $\hat{\Sigma}_U$ and $\tilde{\Sigma}_R$ being the unrestricted and restricted MLE of the residual matrix in the multivariate regression model, respectively.

Specifically, the three tests can be written as:

$$\left.\begin{array}{c} Wald = n\sum_i \lambda_i \\ LM = n\sum_i[\lambda_i/(1+\lambda_i)] \\ \exp(LR) = n\prod_i(1+\lambda_i) \end{array}\right\}.$$

For $H = 3$, there are only two eigenvalues (λ_1 and λ_2), so

$$\left.\begin{array}{c} Wald = \lambda_1+\lambda_2 \\ LM = [\lambda_1/(1+\lambda_1)] + [\lambda_2/(1+\lambda_2)] = \frac{\lambda_1+\lambda_2+2\lambda_1\lambda_2}{\lambda_1+\lambda_2+\lambda_1\lambda_2+1} \\ \exp(LR) = (1+\lambda_1)(1+\lambda_2) = \lambda_1+\lambda_2+\lambda_1\lambda_2+1 \end{array}\right\}.$$

Therefore, the set of values of $Wald$, LM and LR compatible with the previous expressions is a two-dimensional manifold in the three dimensional $(Wald, LM, \exp(LR))$ space, which means the they are not non-linear transformations of each other.

This is in contrast to the case of tests on the coefficients of a multiple regression involving a single $(H = 2)$ regressand or tests on the coefficients of a multivariate regression that involve a single regressor $(K = 1)$, in which case $Wald = \lambda_1$, $LM = \lambda_1/(1 + \lambda_1)$ and $\exp(LR) = (1 + \lambda_1)$, so that all three tests lie on a line (a unidimensional manifold) in the three dimensional $(Wald, LM, \exp(LR))$ space.

3.7.6 *Kickers and Goalkeepers*

Players are divided between kickers and goalkeepers. In brackets is the identification number used in the empirical analysis, and in parentheses it appears the teams they play for.

Kickers [1] Cristiano Ronaldo (Real Madrid/Manchester United), [2] Messi *(Barcelona), [3] Falcao (Atlético de Madrid/Monaco), [4] Gerrard (Liverpool), [5] Guiseppe Rossi (Villareal/Fiorentina), [6] Hulk* (Oporto/Zenit), [7] Ibrahimovic (Inter Milan/Milan/PSG), [8] Kanoute (Sevilla), [9] Negredo* (Almería/Sevilla), [10] Soldado (Getafe/Tottenham), [11] Villa (Valencia/Atlético de Madrid), [12] Xabi Prieto (Real Sociedad), with * denoting the kickers who are left-footed.

Goalkeepers [1] Aouate (Deportivo La Coruña/Mallorca), [2] Diego Alves (Almería/Valencia), [3] Diego López (Villareal/Real Madrid), [4] Iraizoz (Athletic Club Bilbao), [5] Moya (Mallorca/Getafe/), [6] Palop (Sevilla), [7] Ricardo (Osasuna), [8] Roberto (Granada), [9] Ruben (Rayo Vallecano), [10]Tono (Racing Santander/Granada/Rayo Vallecano).

3.7.7 *F approximations*

As I mentioned in section 3.2.2, the Pillai trace test can be written as $V = n^{-1}LM$ while Wilks' lambda is $\Lambda = \exp\left(-n^{-1}LR\right)$ and the Lawley-Hotelling trace test is $LH = n^{-1}Wald$. The F approximations of the Pillai trace (V), Wilks' lambda (Λ) and Lawley-Hotelling (LH) tests that Stata uses are:

$$V_F = \frac{(2n+s+1)V}{(2m+s+1)(s-V)}$$

$$\Lambda_F = \frac{(1-\Lambda^{\frac{1}{t}})df_2}{\left(\Lambda^{\frac{1}{t}}\right)df_1}$$

$$LH_F = \frac{2(sn+1)LH}{s^2(2m+s+1)}$$

where p is the number of columns of y variables, v_h is the hypothesis degrees of freedom, v_e is the error degrees of freedom, $s = \min(p, v_h), m = (|v_h - p| - 1)/2$, $n = (v_e - p - 1)/2$, $df_1 = pv_h$, $df_2 = wt + 1 - pv_h/2$, $w = v_e + v_h - (p + v_h + 1)/2$ and $t = \sqrt{(p^2 v_h^2 - 4)/(p^2 + v_h^2 - 5)}$ (see Stata (2011), Manova, entry for more details).

In addition, Stata uses a degrees of freedom correction $n/(n-K)$ for the heteroskedastitcy robust Wald test in the univariate case (see Stata (2011), Robust, entry for more details).

3.8 Tables Chapter 3

Table 3.2: Distribution of Strategies Combinations and Scoring Rates

	#Obs.	*L-L*	*L-C*	*L-R*	*C-L*	*C-C*	*C-R*	*R-L*	*R-C*	*R-R*
All penalties	549	20.58	2.55	26.78	3.64	1.09	2.37	20.95	0.91	21.13
Scoring rate	86.34	69.91	92.86	97.96	100	0	92.31	95.65	100	78.45

Notes: The first letter refers to the choice made by the kicker (Left (L), Center (C) and Right (R)) and the second one to the choice made by the goalkeeper, always from the point of view of the goalkeeper. For instance, L-R means than the kicker chooses to kick to the left hand side of the goalkeeper and the goalkeeper chooses to jump to his right.

Table 3.3: Test for Equality of Scoring Probabilities with 2 Actions

Player	#Obs.	Frequency L	Frequency R	Scoring Rates L	Scoring Rates R	LM Test	p-value
Kicker 1	44	0.36	0.64	1	0.96	0.58	0.44
Kicker 2	29	0.69	0.31	0.90	0.78	0.78	0.38
Kicker 3	16	0.50	0.50	0.88	0.88	0	1
Kicker 4	32	0.59	0.41	0.95	0.85	0.93	0.33
Kicker 5	21	0.52	0.48	0.91	0.80	0.51	0.48
Kicker 6	22	0.36	0.64	0.79	0.75	0.04	0.85
Kicker 7	41	0.34	0.66	0.93	0.96	0.24	0.63
Kicker 8	9	0.67	0.33	0.67	1	1.29	0.26
Kicker 9	25	0.76	0.24	0.68	1	2.49	0.11
Kicker 10	20	0.25	0.75	1	0.80	1.18	0.28
Kicker 11	20	0.7	0.30	0.93	1	0.45	0.50
Kicker 12	16	0.69	0.31	1	0.80	2.35	0.12
Goalkeeper 1	13	0.62	0.38	0.75	1	1.48	0.22
Goalkeeper 2	16	0.56	0.44	0.56	0.57	0.01	0.95
Goalkeeper 3	10	0.50	0.50	0.60	1	2.50	0.11
Goalkeeper 4	15	0.73	0.27	1	1	0	1
Goalkeeper 5	10	0.10	0.90	1	0.89	0.12	0.73
Goalkeeper 6	10	0.40	0.60	0.50	0.83	1.27	0.26
Goalkeeper 7	9	0.44	0.56	1	0.80	0.90	0.34
Goalkeeper 8	10	0.50	0.50	1	0.80	1.11	0.29
Goalkeeper 9	10	0.60	0.40	0.83	1	0.74	0.39
Goalkeeper 10**	11	0.27	0.73	0.33	1	6.52	0.01

Notes: L (Left) and R (Right) denote the strategies available to the players. Additionally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

Table 3.4: Test for Equality of Scoring Probabilities with 3 Actions

		Frequency			Scoring Rates				
Player	#Obs.	L	C	R	L	C	R	LM Test	p-value
Kicker 1***	50	0.34	0.10	0.56	1	0.60	0.96	11.63	0
Kicker 2	36	0.58	0.11	0.31	0.90	1	0.82	1.11	0.57
Kicker 3	21	0.43	0.19	0.38	0.89	1	0.87	0.53	0.77
Kicker 4**	33	0.58	0.03	0.39	0.95	0	0.85	8.22	0.02
Kicker 6	26	0.54	0.02	0.31	0.79	1	0.75	1.17	0.56
Kicker 7***	42	0.33	0.03	0.64	0.93	0	0.96	13.43	0
Kicker 8	20	0.55	0.20	0.25	0.73	1	1	2.89	0.24
Kicker 9	28	0.68	0.11	0.21	0.68	1	1	3.62	0.16
Kicker 12	20	0.55	0.20	0.25	1	1	0.80	3.16	0.21
Goalkeeper 1	14	0.57	0.07	0.36	0.75	1	1	1.75	0.42
Goalkeeper 2	18	0.50	0.11	0.39	0.56	0.50	0.57	0.03	0.98
Goalkeeper 3	12	0.50	0.08	0.42	0.67	1	0.57	2.40	0.30
Goalkeeper 4**	18	0.67	0.11	0.22	1	0.50	1	8.47	0.01
Goalkeeper 5	15	0.13	0.27	0.60	1	0.85	0.89	0.82	0.66
Goalkeeper 10**	13	0.23	0.08	0.69	0.33	1	1	7.88	0.02

Notes: L (Left), C (Center) and R (Right) denote the strategies available to the players. Additionally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

Table 3.5: Test for Equality of Scoring Probabilities with 6 Actions

Player	#Obs.	Frequency						Scoring Rates						LM Test	p-value
		LD	LU	CD	CU	RD	RU	LD	LU	CD	CU	RD	RU		
Kicker 1***	50	0.22	0.12	0.06	0.04	0.50	0.06	1	1	0.33	1	0.96	1	21.16	0
Kicker 2	36	0.36	0.22	0.03	0.08	0.28	0.03	0.85	1	1	1	0.80	1	2.67	0.75
Kicker 3	21	0.43	-	-	0.19	0.38	-	0.88	-	-	1	0.88	-	0.53	0.77
Kicker 4**	33	0.55	0.03	-	0.03	0.39	-	0.94	1	-	0	0.85	-	8.25	0.04
Kicker 6	26	0.50	0.04	0.08	0.08	0.23	0.08	0.77	1	1	1	0.67	1	2.56	0.77
Kicker 7***	42	0.24	0.10	0.02	-	0.55	0.10	1	0.75	0	-	0.96	1	16.27	0
Kicker 8*	20	0.50	0.05	0.20	-	0.25	-	0.80	0	1	-	1	-	7.45	0.06
Kicker 9	28	0.64	0.04	0.07	0.03	0.21	-	0.67	1	1	1	1	-	4.24	0.37
Kicker 12	20	0.35	0.20	0.10	0.10	0.20	0.05	1	1	1	1	0.75	1	4.21	0.52

Notes: LD (Left Down), LU (Left Up), CD (Center Down), CU (Center Up), RD (Right Down) and RU (Right Up) denote the strategies available to the players. Additionally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

Table 3.6: Test for Serial Independence with 2 Actions

		Transition Matrix					
Player	#Obs.	$L_t \vert L_{t-1}$	$R_t \vert L_{t-1}$	$L_t \vert R_{t-1}$	$R_t \vert R_{t-1}$	LM Test	p-value
Kicker 1	42	0.40	0.60	0.32	0.68	0.27	0.606
Kicker 2	27	0.73	0.27	0.56	0.44	0.92	0.337
Kicker 3	14	0.50	0.50	0.57	0.43	0.08	0.782
Kicker 4	30	0.50	0.50	0.69	0.31	1.15	0.284
Kicker 5	19	0.50	0.50	0.60	0.40	0.20	0.653
Kicker 6	20	0.62	0.38	0.75	0.25	0.40	0.525
Kicker 7	39	0.21	0.79	0.42	0.58	1.74	0.187
Kicker 8	7	0.67	0.33	0.50	0.50	0.18	0.673
Kicker 9	23	0.78	0.22	0.67	0.33	0.27	0.586
Kicker 10	18	0	1	0.29	0.71	1.81	0.179
Kicker 11	18	0.69	0.31	0.83	0.17	0.42	0.516
Kicker 12	14	0.70	0.30	0.80	0.20	0.17	0.680
Goalkeeper 1*	12	0.33	0.67	0.86	0.14	3.75	0.053
Goalkeeper 2	15	0.60	0.40	0.50	0.50	0.15	0.696
Goalkeeper 3	9	0.60	0.40	0.40	0.60	0.40	0.527
Goalkeeper 4*	14	1	0	0.56	0.44	3.64	0.057
Goalkeeper 5	9	0	1	0.13	0.88	0.28	0.598
Goalkeeper 6	9	0.20	0.80	0.60	0.40	1.67	0.197
Goalkeeper 7	8	0.50	0.50	0.33	0.67	0.23	0.635
Goalkeeper 8	9	0.50	0.50	0.50	0.50	0	1
Goalkeeper 9	9	0.75	0.25	0.50	0.50	0.63	0.429
Goalkeeper 10*	10	0	1	0.50	0.50	3.44	0.064

Notes: L_t (Left) and R_t (Right) denote the strategies available to the players at time t while L_{t-1} and R_{t-1} are its corresponding lagged variables. For instance, $L_t|R_{t-1}$ means that the player chooses L at time t after the previous player chose R at $t-1$. Additionally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

Table 3.7: Test for Serial Independence with 3 Actions

		Transition Matrix										
Player	#Obs.	$L_t \| L_{t-1}$	$C_t \| L_{t-1}$	$R_t \| L_{t-1}$	$L_t \| C_{t-1}$	$C_t \| C_{t-1}$	$R_t \| C_{t-1}$	$L_t \| R_{t-1}$	$C_t \| R_{t-1}$	$R_t \| R_{t-1}$	LM Test	p-value
Kicker 1	49	0.47	0	0.53	0.60	0	0.40	0.21	0.18	0.61	1.89	0.12
Kicker 2	34	0.60	0.15	0.25	0.75	0	0.25	0.45	0.10	0.45	0.53	0.71
Kicker 3	20	0.40	0.10	0.50	0.67	0	0.33	0.38	0.38	0.24	0.86	0.50
Kicker 6	25	0.50	0.07	0.43	0.50	0	0.50	0.63	0.37	0	1.95	0.12
Kicker 7	41	0.20	0.07	0.73	0	0	1	0.42	0	0.58	1.01	0.41
Kicker 8	19	0.41	0.18	0.41	0.75	0.25	0	0.75	0.25	0	1.07	0.39
Kicker 9	27	0.60	0.10	0.30	1	0	0	0.83	0.17	0	0.84	0.51
Kicker 12	19	0.42	0.25	0.33	1	0	0	0.60	0.20	0.20	0.79	0.54
Goalie 1	13	0.62	0	0.38	1	0	0	0.40	0.20	0.40	0.61	0.66
Goalie 2	17	0.56	0	0.44	0.50	0	0.50	0.42	0.29	0.29	0.83	0.51
Goalie 3	11	0.33	0.17	0.50	0	0	1	0.80	0	0.20	0.96	0.45
Goalie 4	17	0.73	0.09	0.18	0.50	0.50	0	0.60	0	0.40	1.13	0.36
Goalie 5	14	0.50	0	0.50	0	0.60	0.40	0.12	0.12	0.76	1.68	0.19
Goalie 10	12	0	0	1	0	0	1	0.30	0.10	0.60	0.36	0.84

Notes: L_t (Left), C_t (Center) and R_t (Right) denote the strategies available to the players at time t while L_{t-1}, C_{t-1} and R_{t-1} are its corresponding lagged variables. For instance, $L_t|R_{t-1}$ means that the player chooses L at time t after the previous player chose R at $t-1$. Additionally, * indicates rejection of the null at the 10% significance level, ** 5% level, *** 1% level.

References

Anatolyev, S. and G. Kosenok (2009): "Tests in Contingency Tables as Regression Tests", Economics Letters, 105, pp. 189–192.

Apesteguia, J. and I. Palacios-Huerta (2010): "Psychological Pressure in Competitive Environments: Evidence from a Randomized Natural Experiment", American Economic Review, 100, pp. 2548-64.

Aradillas-López, A., A. Gandhi and D. Quint (2013): "Identification and Inference in Ascending Auctions with Correlated Private Values", Econometrica, 81, pp. 489-534.

Athey, S. and P. A. Haile (2002): "Identification of Standard Auction Models", Econometrica, 70, pp. 2107-2140.

Athey, S., J. Levin and E. Seira (2011): "Comparing Open and Sealed Bid Auctions: Evidence from Timber Auctions", Quarterly Journal of Economics, 126, pp. 207-257.

Bajari, P. and A. Hortaçsu (2003): "The Winner's Curse, Reserve Prices and Endogenous Entry: Empirical Insights from eBay Auctions", RAND Journal of Economics, 34, pp. 329-355.

Balzer, B. and A. Rosato (2020): "Expectations-Based Loss Aversion in Auctions with Interdependent Values: Extensive vs. Intensive Risk", Management Science.

Banerji, A. and N. Gupta (2014): "Detection, Identification, and Estimation of Loss Aversion: Evidence from an Auction Experiment", American Economic Journal: Microeconomics, 6, pp. 91-133.

Barberis, N., M. Huang and T. Santos (2001): "Prospect Theory and Asset Prices", Quarterly Journal of Economics, 116, pp. 1-53.

Barberis, N. and M. Huang (2008): "Stocks as Lotteries: The Implications of Probability Weighting for Security Prices", American Economic Review, 98, pp. 2066-2100.

Belot, M., V. Bhaskar and J. van de Ven (2010): "Promises and Cooperation: Evidence from a TV Game Show", Journal of Economic Behavior & Organization, 73, pp. 396-405.

Belot, M., V.P. Crawford and C. Heyes (2013): "Players of Matching Pennies Automatically Imitate Opponents' Gestures Against Strong Incentives", Proceedings of the National Academy of Sciences, 110, pp. 2763-2768.

Berndt, E.R. and N.E. Savin (1977): "Conflict Among Criteria for Testing Hypotheses in the Multivariate Linear Regression Model", Econometrica, 45, pp. 1263-1277.

Bouissou, M. B., J. Laffont and Q. H. Vuong (1986): "Tests of Noncausality Under Markov Assumptions for Qualitative Panel Data", Econometrica, 54, pp. 395-414.

Bradley, J. V. (1968): "Distribution-Free Statistical Tests", First Edition, Prentice-Hall.

Brown, J. and R. Rosenthal (1990): "Testing the Minimax Hypothesis: A Reexamination of O'Neill's Experiment", Econometrica, 58, pp. 1065-1081.

Cameron, A. C. and P. K. Trivedi (2005): Microeconometrics: Methods and Applications, Cambridge University Press.

Chiappori, P-A, S. Levitt and T. Groseclose (2002): "Testing Mixed-Strategy Equilibria When Players Are Heterogeneous: The Case of Penalty Kicks in Soccer", American Economic Review, 92, pp. 1138-1151.

Dastoor, N. K. (2001): "Conflict Among Criteria for Testing Hypotheses: Examples from Non-Normal Distributions", Econometric Theory, 17, pp. 487-491.

Davidson, R. and J. MacKinnon (1998): "Graphical Methods for Investigating the Size and Power of Hypothesis Tests", The Manchester School, 66, pp. 1-26.

Dionne, G., P. St-Amour and D. Vencatachellum (2009): "Asymmetric Information and Adverse Selection in Mauritanian Slave Auctions", Review of Economic Studies, 76, pp. 1269-1295.

Dodonova, A. and Y. Khoroshilov (2005): "English Auctions with Loss Averse Bidders", Working paper, University of Ottawa Telfer School of Management.

Dodonova, A. and Y. Khoroshilov (2009): "Behavioral Biases in Auctions: An Experimental Study", Economics Bulletin, 29, pp. 2223-2231.

Dridi, R., A. Guay and E. Renault (2007): "Indirect Inference and Calibration of Dynamic Stochastic General Equilibrium Models", Journal of Econometrics, 136, pp. 397-430.

Eisenhuth, R. and M. Grunewald (2020): "Auctions with Loss-averse Bidders", International Journal of Economic Theory, 16, pp. 129-152.

Engle, R.F (1983): "Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics", Handbook of Econometrics, vol. 2, eds. Z. Griliches and M.D. Intriligator, chapter 13, pp. 775-826.

Erev, I. and A. E. Roth (1998): "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria", American Economic Review, 88, pp. 848–881.

Federation Internationale de Football Association (2018): "Official Laws of the Game", www.fifa.com/development/education-and-technical/referees/laws-of-the-game.html

Fisher, R. A (1922): "On the Interpretation of $\mathcal{X}^2$ from Contingency Tables, and the Calculation of P", Journal of the Royal Statistical Society, 85, pp. 87-94.

Garicano, L., I. Palacios-Huerta and C. Prendergast (2005): "Favoritism Under Social Pressure", Review of Economics and Statistics, 87, pp. 208-216.

Gentry, M. and T. Li (2014): "Identification in Auctions with Selective Entry", Econometrica, 82, pp.315-344.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators", Econometrica, 50, pp. 1029-1054.

Henriksson, R.D. and R.C. Merton (1981): "On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills", Journal of Business, 54, pp. 513-533.

Hong, H. and M. Shum (2002): "Increasing Competition and the Winner's Curse: Evidence from Procurement", Review of Economic Studies, 69, pp. 871-898.

Hong, H. and M. Shum (2003): "Econometric Models of Asymmetric Ascending Auctions", Journal of Econometrics, 112, pp. 327-358.

Hsu, S-H., C-Y. Huang, and C-T. Tang (2007): "Minimax Play at Wimbledon: Comment", American Economic Review, 97, pp. 517-523.

Judge, G. , W. Griffiths, H. Lütkepohl, T-C. Lee and R. Carter Hill (1985): The Theory and Practice of Econometrics, Wiley, Second edition.

Kagel, J. H. and D. Levin (1986): "The Winner's Curse and Public Information in Common Value Auctions", American Economic Review, 76, pp. 894-920.

Kahneman, D. and A. Tversky (1979): "Prospect Theory: An Analysis of Decision Under Risk", Econometrica, 47, pp. 263-291.

Koptyug, N. (2016): "Asymmetric Information in Auctions: Are Resellers Better Appraisers?", Working Paper, Research Institute of Industrial Economics (IFN).

Kőszegi, B. and M. Rabin (2006): "A Model of Reference-Dependent Preferences", Quarterly Journal of Economics, 121, pp. 1133-1165.

Krasnokutskaya, E. and K. Seim (2011): "Bid Preference Programs and Participation in Highway Procurement Auctions", American Economic Review, 101, pp. 2653-86.

Laffont, J.J., H. Ossard and Q. Vuong (1995): "Econometrics of First-price Auctions", Econometrica, 63, pp. 953-980.

Lange, A. and A. Ratan (2010): "Multi-dimensional Reference-dependent Preferences in Sealed-bid Auctions–How (Most) Laboratory Experiments Differ from the Field", Games and Economic Behavior, 68, pp. 634-645.

Lee, T., G. Judge and A. Zellner (1968): "Maximum Likelihood and Bayesian Estimation of Transition Probabilities", Journal of the American, Statistical Association, 63, pp. 1162-1179.

Levitt, S., J. List and D. Reiley (2010): "What Happens in the Field Stays in the Field: Professionals Do Not Play Minimax in Laboratory Experiments", Econometrica, 78, pp. 1413-1434.

List, J.A. (2004): "Neoclassical Theory Versus Prospect Theory: Evidence from the Marketplace", Econometrica, 72, pp. 615-625.

List, J.A. (2006): "Friend or Foe? A Natural Experiment of the Prisoner's Dilemma", Review of Economics and Statistics, 88, pp. 463-471.

Magnus, J. (2007): "The Asymptotic Variance of the Pseudo Maximum Likelihood Estimator", Econometric Theory, 23, pp. 1022–1032.

Magnus, J. and H. Neudecker (1988): Matrix Differential Calculus with Applications in Statistics and Econometrics, Wiley.

McFadden, D.L. (1996): "Lectures on Simulation-Assisted Statistical Inference," Working Paper, University of California–Berkeley.

Mellström, C. and M. Johannesson (2008): "Crowding Out in Blood Donation: Was Titmuss Right?", Journal of the European Economic Association, 4, pp. 845-863.

Miklós-Thal, J. and H. Ullrich (2016): "Career Prospects and Effort Incentives: Evidence from Professional Soccer", Management Science, 62, pp. 1645-1667.

Milgrom, P. and R.J. Weber (1982): "A Theory of Competiting Bidding", Econometrica, 50, pp. 197-216.

Miller, C. (1998): He Always Puts It to the Right. A concise History of the Penalty Kick, Edmundsbury Press.

Mood, A., F-A. Graybill and D. Boes (1974): Introduction to the Theory of Statistics, Third edition, McGraw-Hill.

Newey, W.K. and K.D. West (1987):. "Hypothesis Testing with Efficient Method of Moments

Estimation", International Economic Review, 28, pp. 777-787.

O'Neill, B. (1987): "Nonmetric Tests of the Minimax Theory of Two Person Zero Sum Games", Proceedings of the National Academy of Sciences, 84, pp. 2106-2109.

Osborne, J. (2003): An Introduction to Game Theory, Oxford University Press, First edition.

Paarsch, H.J. (1997): "Deriving an Estimate of the Optimal Reserve Price: an Application to British Columbian Timber Sales", Journal of Econometrics, 78, pp. 333-357.

Palacios-Huerta, I.(2003): "Professionals Play Minimax", Review of Economic Studies, 70, pp. 395-415.

Palacios-Huerta, I. (2017): "Strictly Competitive Strategic Situations", Economia Industrial, 403, pp. 19-28.

Palacios-Huerta, I. and O. Volij. (2008): "Experientia Docet: Professionals Play Minimax in Laboratory Experiments", Econometrica, 76, pp. 71-115.

Pesaran, M.H and A. Timmermann (1994): "A Generalization of the Non-Parametric Henriksson-Merton Test of Market Timing", Economics Letters, 44, pp. 1-7.

Pope, D.G. and M.E. Schweitzer (2011): "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes", American Economic Review, 101, pp. 129-57.

Post, T., M.J. van den Assem, G. Baltussen and R.H. Thaler (2008): "Deal or no Deal? Decision Making Under Risk in a Large-Payoff Game Show", American Economic Review, 98, pp. 38-71.

Rosato, A. and A.A. Tymula (2019): "Loss Aversion and Competition in Vickrey Auctions: Money Ain't No Good", Games and Economic Behavior, 115, pp. 188-208.

Roth, A.E. and A. Ockenfels (2002): "Last-minute Bidding and the Rules for Ending Second-price Auctions: Evidence from eBay and Amazon Auctions on the Internet", American Economic Review, 92, pp. 1093-1103.

Ruud, P. (2000): An Introduction to Classical Econometric Theory, Oxford University Press.

Shachat, J., J.T. Swarthout and L. Wei (2015): "A Hidden Markov Model for the Detection of Pure and Mixed Strategy Play in Games", Econometric Theory, 31, pp. 729-752.

Sherman, J. and W.J. Morrison (1950): "Adjustment of an Inverse Matrix Corresponding to a Change in One Element of a Given Matrix", Annals of Mathematical Statistics, 21, pp. 124-127.

Song, U. (2004): “Nonparametric Estimation of an E-Bay Auction Model with an Unknown Number of Bidders,” Working paper, University of British Columbia.

Sprenger, C. (2015): "An Endowment Effect for Risk: Experimental Tests of Stochastic Reference Points", Journal of Political Economy, 123, pp. 1456-1499.

StataCorp. (2011): Stata 12 Base Reference Manual, Stata Press, College Station, TX.

Stewart, K.G. (1995): "The Functional Equivalence of the W, LR and LM statistics", Economic Letters, 49, pp. 109-112.

Swanson, N.R. and H. White (1997): "A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks", Review of Economics and Statistics, 79, pp. 540-550.

Tversky, A. and D. Kahneman (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty", Journal of Risk and Uncertainty, 5, pp. 297-323.

van Dolder, D., M.J. van den Assem, C.F. Camerer and R.H. Thaler (2015): "Standing United or Falling Divided? High Stakes Bargaining in a TV Game Show", American Economic Review, 105, pp. 402-07.

Vickrey, W. (1961): "Counterspeculation, Auctions, and Competitive Sealed Tenders", Journal of Finance, 16, pp. 8-37.

von Wangenheim, J. (2017): "English Versus Vickrey Auctions with Loss Averse Bidders", Working paper, Humboldt University Berlin.

Walker, M. and J. Wooders (2001): "Minimax Play at Wimbledon", American Economic Review, 91, pp. 1521-1538.

Wilson, R. (1998): "Sequential Equilibria of Asymmetric Ascending Auctions", Economic Theory, 12, pp. 433-440.

Wooldridge, J. (2002): Introductory Econometrics: A Modern Approach, Second Edition, South-Western Econometrics.

Zaninetti, L. (2017): "A Left and Right Truncated Lognormal Distribution for the Stars", Advances in Astrophysics 2, 197.