# Essays on Econometrics

by

Zhengfei Yu

B.A., Zhejiang University, 2009
M.A., The University of British Columbia, 2010

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Economics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

July 2015

# Abstract

This thesis studies two topics in Econometric models, multiple equilibria and weak instruments. Chapter 1 is an introduction.

Chapter 2 considers nonparametric structural equations which may have multiple solutions for the endogenous variables. The main finding is that multiple equilibria would reveal itself in the form of jump(s) in the density function of the endogenous variables. When there is a unique equilibrium, the density function of endogenous variables will be continuous, while when there are multiple equilibria, the density will have a jump at some point, under reasonable conditions. Our test statistic is based on maximizing local jumps over the support of endogenous variables and the critical value is computed via a Gaussian multiplier bootstrap.

Chapter 3 shows that in games with incomplete information, even when the payoff functions and the latent distributions are all smooth, the observed conditional choice probabilities may have a jump with respect to continuous covariates. This chapter provides a theoretical analysis on the relationship between the equilibrium behaviour of the game and the presence of a jump in the conditional choice probabilities. Such jump(s) matters in empirical research for two reasons. Statistically, it affects the estimation of the conditional choice probabilities. Economically, whether the conditional choice probabilities have a jump or not reveals information about the equilibrium behaviour of the game. Our findings are robust to correlated private information and unobserved heterogeneity independent of covariates.

Chapter 4 considers efficient inference for the coefficient of the endogenous variable in linear regression models with weak instrumental variables (Weak-IV). We focus on the power of tests for the alternative hypotheses that are determined by arbitrarily large deviations from the null. We derive the power envelope for such alternatives in the Weak-IV scenario. Then we compare the power properties of popular Weak-IV robust tests, focusing on the Anderson-Rubin (AR) and Conditional Likelihood Ratio (CLR) tests. We find that their relative performance depends on the degree of endogeniety in the model. In addition, we propose a Conditional Lagrange Multiplier (CLM) test. We also extend our analysis to heteroskedastic models.

# Preface

Chapter 4 is co-authored with Prof. Vadim Marmer. I have been actively participating in all stages of this project, including reviewing the literature, deriving and proving propositions, conducting numerical computation, and writing the manuscript.

# Table of Contents


**Abstract** . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . ii

**Preface** . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . iv

**Table of Contents** . . . . . . . . . . . . . . . . . . . . . . . . . v

**List of Tables** . . . . . . . . . . . . . . . . . . . . . . . . . . viii

**List of Figures** . . . . . . . . . . . . . . . . . . . . . . . . . . ix

**Acknowledgements** . . . . . . . . . . . . . . . . . . . . . . . . . x

**1 Introduction** . . . . . . . . . . . . . . . . . . . . . . . . . . . 1
1.1 Dissertation Outline . . . . . . . . . . . . . . . . . . . . . . . 2

**2 Testing for Multiple Equilibria in Continuous Dependent Variables** . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
2.1.1 Related Literature . . . . . . . . . . . . . . . . . . . . . . 7
2.1.2 Organization of Chapter 2 . . . . . . . . . . . . . . . . . . 9
2.2 The Framework and Examples . . . . . . . . . . . . . . . . . . 9
2.2.1 The Econometric Model . . . . . . . . . . . . . . . . . . . 9
2.2.2 Examples . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
2.3 Testing the Presence of Multiple Equilibria via Discontinuity 13
2.3.1 An Equilibrium Selection Rule . . . . . . . . . . . . . . . 13
2.3.2 A Testing Criterion . . . . . . . . . . . . . . . . . . . . . 18
2.3.3 An Extension: Jumps in the Latent Distribution . . . 30
2.4 A Test Statistic . . . . . . . . . . . . . . . . . . . . . . . . . . 31
2.4.1 Construction of the Test . . . . . . . . . . . . . . . . . . 31
2.4.2 Asymptotic Properties of the Test . . . . . . . . . . . . 36
2.5 Monte Carlo Simulations . . . . . . . . . . . . . . . . . . . . . 38
2.5.1 Data Generating Processes (DGP) . . . . . . . . . . . . 38

2.5.2 Performance of the Test . . . . . . . . . . . . . . . . . 43
2.6 Conclusions and Remarks . . . . . . . . . . . . . . . . . . . . 47

**3 Jumps of the Conditional Choice Probabilities in Incomplete Information Games** . . . . . . . . . . . . . . . . . . . . . . . . 49
3.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . 49
3.1.1 Related Literature . . . . . . . . . . . . . . . . . . . . 50
3.1.2 Organization of Chapter 3 . . . . . . . . . . . . . . . 52
3.2 The Set-up of the Game . . . . . . . . . . . . . . . . . . . . 52
3.3 An Implication of Equilibrium Behaviour on Data . . . . . . 55
3.3.1 The Equilibrium Behaviour and Presence of A Jump in Conditional Choice Probabilities . . . . . . . . . . 57
3.3.2 A Numerical Example . . . . . . . . . . . . . . . . . . 65
3.3.3 An Extension: Unobserved Heterogeneity . . . . . . . 67
3.3.4 Testing for Multiple Equilibria via Discontinuity . . . 73
3.4 An Estimation Problem Due to Jumps . . . . . . . . . . . . 75
3.5 Testing for the Presence of Jump(s) . . . . . . . . . . . . . 77
3.6 Conclusions and Remarks . . . . . . . . . . . . . . . . . . . . 79

**4 Efficient Inference in Econometric Models When Identification Can Be Weak** . . . . . . . . . . . . . . . . . . . . . . . . . 81
4.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . 81
4.1.1 Organization of Chapter 4 . . . . . . . . . . . . . . . 83
4.2 An Asymptotic Experiment for Linear IV Models . . . . . . 83
4.3 The Optimal Rotational Invariant and Asymptotically Similar Test . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 87
4.4 The Power Envelope Under an Unknown Nuisance Parameter 91
4.5 Power Comparisons of Robust Tests . . . . . . . . . . . . . . 94
4.5.1 Popular Weak-IV Robust Tests . . . . . . . . . . . . 94
4.5.2 A New Test: Conditional Lagrange Multiplier (CLM) Test . . . . . . . . . . . . . . . . . . . . . . . . . . . 96
4.5.3 Power Calculations . . . . . . . . . . . . . . . . . . . 100
4.6 Heteroskedastic Models . . . . . . . . . . . . . . . . . . . . . 107
4.6.1 The Optimal Rotational Invariant and Asymptotically Similar Test In Heteroskedastic Models . . . . . . . . 110
4.6.2 The Generalized Likelihood Ratio (GLR) Statistic In Heteroskedastic Models . . . . . . . . . . . . . . . . . 112
4.7 Concluding Remarks . . . . . . . . . . . . . . . . . . . . . . 115

**Bibliography** . . . . . . . . . . . . . . . . . . . . . . . . . . . . 116

**Appendices** . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

**A Appendix for Chapter 2** . . . . . . . . . . . . . . . . . . . . . . . . 121
A.1 Proofs for Section 2.3 . . . . . . . . . . . . . . . . . . . . . . . . 121
A.2 Conditions in Section 2.4 . . . . . . . . . . . . . . . . . . . . . . 131
A.3 Proofs for Section 2.4 . . . . . . . . . . . . . . . . . . . . . . . . 133

**B Appendix for Chapter 3** . . . . . . . . . . . . . . . . . . . . . . . . 153
B.1 A Discussion of the Alternative Equilibrium Notion . . . . . . 153
B.2 Mathematical Proofs . . . . . . . . . . . . . . . . . . . . . . . . 154

# List of Tables


2.1 Conditions for multiple equilibria to produce jump(s) in density 22
2.2 Equilibrium prices . . . . . . . . . . . . . . . . . . . . . . . . 41
2.3 Rejection frequencies of the test . . . . . . . . . . . . . . . . . . 45
2.4 Rejection frequencies under different bandwidths . . . . . . . 46
2.5 Rejection frequencies under different maximizing sets . . . . . 47

3.1 Normal-form of the game . . . . . . . . . . . . . . . . . . . . . 53
3.2 Normal-form of the numerical example . . . . . . . . . . . . . 65
3.3 Parameters for the games . . . . . . . . . . . . . . . . . . . . . 67
3.4 Normal-form of the game with $T = t$ . . . . . . . . . . . . . . . 68

# List of Figures


2.1 Illustration of jumps in density . . . . . . . . . . . . . . . . . . 6
2.2 Reduced-form functions . . . . . . . . . . . . . . . . . . . . . . . 15
2.3 Multiple equilibria and jump(s) . . . . . . . . . . . . . . . . . . 19
2.4 Characterization of equilibria . . . . . . . . . . . . . . . . . . . 40
2.5 Data generated by DGP0,1,2,3 . . . . . . . . . . . . . . . . . . 42
2.6 Jump location curves . . . . . . . . . . . . . . . . . . . . . . . . 44

3.1 The equilibrium belief(s) $\sigma_1$ . . . . . . . . . . . . . . . . . . . . 66
3.2 Equilibria correspondence from covariate $x$ to belief $\sigma_1$ . . . 68
3.3 Observed conditional choice probability . . . . . . . . . . . . 69

4.1 Power comparisons among robust tests, 2 IVs . . . . . . . . . 101
4.2 Power comparisons among robust tests, 5 IVs . . . . . . . . . 102
4.3 Relative power of the AR versus CLR tests, 2 IVs . . . . . . 103
4.4 Relative power of the AR versus CLR tests, 5 IVs . . . . . . 104
4.5 Relative power of the AR versus CLR tests, 10 IVs . . . . . . 105
4.6 Relative power of the AR versus CLR tests, 20 IVs . . . . . . 106

# Acknowledgements

I am very grateful to my thesis supervisors, Prof. Vadim Marmer and Prof. Kyungchul (Kevin) Song, for their enlightening guidance, helpful advice, and continuous encouragement.

I would like to thank Prof. Hiroyuki Kasahara for his valuable comments and generous support. I also benefited a lot from Prof. Yanqin Fan and Prof. Paul Schrimpf.

Special thanks are owed to my parents, who have supported me throughout my years of pursuing this degree. They are thousands of miles away from their only child, but they always pass their confidence and optimism to me.

I thank my friend Zhe Chen for living through the Ph.D life together with me, thank my colleague Jun Ma from whom I have learned a lot of econometrics.

Thanks also go to my old friends, Zi (Lindsay) Lin, Hui Jiang, Tianfang Wu and Yike Mao. Although I have experienced a lot of “new” things in the past ten years, fortunately I still keep the curiosity, enthusiasm and belief that we ever shared.

# Chapter 1

# Introduction

This thesis studies two topics in econometric models: multiple equilibria and weak instruments.

The presence of multiple equilibria poses challenges for comparative static and conterfactual experiments (see Berry, Levinsohn and Pakes (1999), Echenique and Komunjer (2009) and Borkovsky et al. (2014)), and it may also effect the identification and estimation approaches (see Aradillas-Lopez (2010), Paula (2012), Aguirregabiria and Mira (2013), Wan and Xu (2014) and Berry and Haile (2014), among others). Chapter 2 considers nonparametric structural equations which may admit multiple solutions for the dependent variables. The main finding is that uniqueness or multiplicity of equilibria produces testable implications on the continuity or discontinuity of the density function of dependent variables. Then we propose a test for multiple equilibria based on checking the existence of jump(s) in the density function of dependent variables. Chapter 3 considers incomplete information games with possibly correlated private information. We show that even when the payoff functions and the latent distributions are all smooth, the observed conditional choice probabilities may have jump(s) with respect to the continuous covariates. Statistically, the possibility of such a jump affects estimation of the conditional choice probabilities. Economically, the presence of such jumps reveals information about equilibrium behaviour of the game.

Weak instrumental variables have received a lot of attention in econometrics (see Staiger and Stock (1997), Kleibergen (2002, 2007), Moreira (2001, 2003) and Andrews, Moreira and Stock (2006)). Chapter 4 studies efficient inference of the structural parameter in linear instrumental variables regression models when the instruments may be weak, with a focus on the alternatives that are determined by arbitrarily large deviations from the null.

## 1.1 Dissertation Outline

Chapter 2, *Testing for Multiple Equilibria in Continuous Dependent Variables*, proposes a test for the presence of multiple equilibria when the structural equations are nonparametric and the dependent variables are continuous. Multiple equilibria may arise when the structural equations admit multiple solutions. Such multiplicity can happen in economic models of price competitions, social interactions, macroeconomics and many other fields. This chapter finds that multiple equilibria would reveal itself in the form of discontinuities in the (conditional) density function of the dependent variables. Under some regularity assumptions, a model with a unique equilibrium necessarily makes the conditional density continuous with respect to the endogenous variables. On the other hand, under reasonable conditions, a model with multiple equilibria produces jump(s) in the conditional density with respect to the endogenous variables. Such jump(s) is typically caused by changes in the number of solutions to the structural equations. With an additional assumption, uniqueness or multiplicity of equilibrium leads to continuity or discontinuity in the unconditional density function of endogenous variables. This way we transform the problem of testing for multiple equilibria into testing for the presence of a jump in the density of dependent variables. Our testing procedure consists of three steps. Firstly, for a given point, we compute the local jump of the density function. Secondly, we obtain the test statistic by computing the maximal local jump. Lastly, we simulate the critical value via the Gaussian multiplier bootstrap proposed by Chernozhukov, Chetverikov and Kato (2013).

In Chapter 3 titled *Jumps of the Conditional Choice Probabilities in Incomplete Information Games*, we show that in static incomplete information games, the conditional choice probabilities may have a jump with respect to the continuous covariates, even when the payoff functions and latent distributions are all smooth. The possibility of such jump(s) would affect the estimation strategy of the conditional choice probabilities. We further establish the relationship between the equilibrium behaviour and jump(s) in the conditional choice probabilities. In particular, when the equilibrium characterizing equations always admit a unique solution, the conditional choice probabilities will be continuous. On the other hand, when there are multiple equilibria, or the single equilibrium present in the data varies in types, the conditional choice probabilities will have a jump, except for some special equilibrium selection rules. Such a relationship is robust to correlated private information across players and to unobserved heterogeneity independent of covariates. Hence testing for the presence of a jump in the conditional choice probabilities also provides information about the equilibrium behaviour of the game.

The last chapter, *Efficient Inference in Econometric Models When Identification Can Be Weak*, considers efficient inference for the coefficient of the endogenous variable in instrumental variables regression models with weak instrumental variables (Weak-IV). We focus on the power of tests for alternatives that are determined by arbitrarily large deviations from the null. We derive the power envelope for such alternatives in the Weak-IV scenario. Then we compare the power properties of popular Weak-IV robust tests, focusing on the Anderson-Rubin (AR) and Conditional Likelihood Ratio (CLR) tests. We find that their relative performance depends on the degree of endogeniety in the model. This is different from Andrews, Moreira and Stock (2006), which found that the CLR test numerically dominates the AR test when the weighted average power is concerned. In addition, we propose the Conditional Lagrange Multiplier (CLM) test, which is asymptotically efficient when the instruments are strong, robust to Weak-IV, and exhibits the same power as the AR test for arbitrarily large deviations from the null. We also study the heteroskedastic case, and find that the generalized likelihood ratio statistic under heteroskedasticity reduces to the AR statistic in the Weak-IV scenario and when the alternatives are determined by arbitrarily large deviations from the null.

# Chapter 2

# Testing for Multiple Equilibria in Continuous Dependent Variables

## 2.1 Introduction

Multiple equilibria are commonly generated by economic models in industrial organization, social interactions, macroeconomics, and many other fields. From an econometric perspective, multiple equilibria can be regarded as situations in which the exogenous variables cannot uniquely determine the endogenous variables[1]. Jovanovic (1989) described three cases where multiple equilibria may arise: optimization problems, linear simultaneous equations and non-linear simultaneous equations.

Detecting the presence of multiple equilibria matters for several reasons. Firstly, multiplicity of equilibria poses challenges for comparative statics and counterfactual analysis. Empirical researchers usually assume the uniqueness of equilibrium when conducting counterfactual experiments (see for example, Berry, Levinsohn and Pakes (1999).[2]). When multiple equilibria arise, the predicted outcome from a policy change becomes indeterminate. Secondly, uniqueness or multiplicity of equilibria is closely related to identification of nonparametric simultaneous equations. While nonparametric simultaneous equations can be identified under an abstract completeness conditions, an alternative identification approach requires uniqueness of equilibrium as a maintained assumption for identification (see Berry and Haile (2013, 2014), Mazkin (2008), among others for identification of nonparametric simultaneous equations, and Horowitz (1996) for the transformation model). Thirdly, we will show that

[1] We use dependent variables, endogenous variables and outcome variables interchangeably in this chapter.

[2] They wrote: "This assumes both that the equilibrium without the VER is also Nash in prices and that the equilibrium is unique (or at least that we solve for the relevant one)."

multiplicity of equilibria is associated with jump(s) in the density function of the dependent variables, under regularity conditions. This will affect the estimation approach for the density. Lastly, knowing the presence of multiple equilibria helps understanding whether the variation in the outcomes can be better explained by fundamentals or a self-fulling mechanism.[3]

While tests for multiple equilibria in discrete games have been developed by Paula and Tang (2012) and Aguirregabiria and Mira (2013), among others, research for continuous dependent variables (for example, prices) with multiple equilibria has been rare. To the best of my knowledge, this chapter makes the first attempt to propose a test for uniqueness against multiplicity of equilibria when the dependent variables are continuous and the model is nonparametric. Our framework is a system of equations $r(Y) = g(X) + U$, where $r(\cdot)$ and $g(\cdot)$ are not parametrically specified. The observed data is an i.i.d. sample $\{X_i, Y_i\}_{i=1}^n$. Here we do not make the assumption that the unknown function $r(\cdot) : \mathcal{Y} \to \mathbf{R}^J$ is one to one, where $\mathcal{Y} \subset \mathbf{R}^J$. Therefore, the dependent variables $Y$ may not be uniquely determined by exogenous variables $(X, U)$. An equilibrium selection rule is introduced to complete the model. Such a framework is general enough to include applications varying from price competitions to social interactions.

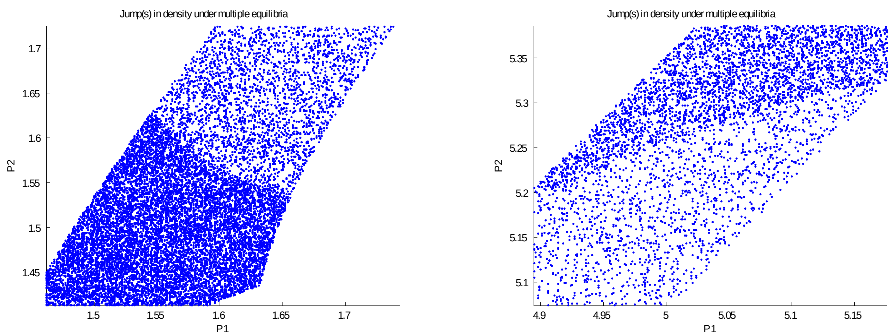
For structural equations $r(y) = g(x) + u$ and an equilibrium selection rule, we say there is a unique equilibrium if the structural equations $r(y) = g(x) + u$ admit a unique solution for $y$ given any value $(x, u)$, or they may admit multiple solutions, but the equilibrium to be present in the data is uniquely determined by the value $(x, u)$. In contrast, we say there are multiple equilibria if the structural equations admit multiple solutions at some $(x, u)$, and more than one of the solutions can be realized in the data with positive probabilities, after fixing $(x, u)$. Assuming that the functions $r$ and $g$ are twice continuously differentiable, unobservable $U$ has a continuous density function, and $U$ is independent of $X$, this chapter finds that when there is a unique equilibrium, the conditional density $f_{Y|X}(y|x)$ is continuous[4] in $y$ within its support, for all $x$. On the other hand, when there are multiple equilibria, under reasonable conditions, the conditional density $f_{Y|X}(y|x)$ will have a jump in $y$, such that the limits

[3]For example, Dagsvik and Javanovic (1994) considered the question "was the Great Depression the outcome of a massive coordination failure? Or was it a unique equilibrium respond to adverse shocks?"

[4]Throughout this chapter, continuity at a point means that the limits from all direction coincide. It does not place any condition on the value of the function at the limit point.

Figure 2.1: Illustration of jumps in density

Jumps in the lower prices Jumps in the higher prices




from different directions are not all equal.[5] With an additional assumption, uniqueness or multiplicity of equilibria leads to continuous or discontinuous unconditional density function $f_Y(y)$. The source of such jump(s) can be described as follows. The value of dependent variables are determined in equilibrium. Thus the (conditional) density of dependent variables depends on an equilibrium selection rule, which picks one from multiple solutions with a selection probability. When there are multiple equilibria, one would typically see that the number of solutions to the structural equations varies as explained below. Suppose that there is a region where each $y$ is realized as the unique solution to the structural equations, while there is another region where each $y$ is realized as one of multiple solutions. At one side of the boundary between above two regions, the selection probability is always 1, while on the other side of the boundary, the selection probability tends to be below 1 because now the selection rule has to pick one equilibrium from multiple candidates. Thus the selection probability has a jump between regions with distinct numbers of solutions. This would create a jump in the density of dependent variables.

Figure 2.1 shows such jumps in the density function of data $(P_1, P_2)$ generated by a simplified Berry, Levinsohn and Pakes (BLP) model but with heteroskedasticity in the idiosyncratic term of the consumers' preference. The details of the model setup and parameter choices are given in Section 2.5. Here the dependent variables are prices $P_1$ and $P_2$. We can clearly see that the thickness of the points has an abrupt change on a line: this is the

[5]Throughout this chapter, a discontinuity or a jump refers to a jump discontinuity where the directional limits are not all equal.

jump location curve of the joint density of prices, $f_{P_1,P_2}(p_1, p_2)$.

This way we translate the problem of testing for multiple equilibria to testing for the presence of a jump in the density function. Our testing procedure consists of three steps. Firstly, we compute the local jump of the density at a fixed $y$. In doing this, we check directions along all coordinates of $y$. Secondly, we maximize those local jumps over the support of $Y$. This gives our test statistic. Lastly, we compute the critical value via the Gaussian multiplier bootstrap proposed by Chernozhukov, Chetverikov and Kato (2013).

### 2.1.1 Related Literature

In a discrete game with incomplete information and independent types of players, tests for multiple equilibria have been developed by Paula and Tang (2012) and Aguirregabiria and Mira (2013). Their tests were based on dependence among players' choices. We use a different testing criterion, which fits our framework of continuous dependent variables. Besides, our framework is nonparametric in both structural equations and the equilibrium selection rule. This is different from Dagsvik and Jovanovic (1994), which formulated a parametric model and a parametric equilibrium selection rule to study whether depression can be explained as a low level equilibrium. Our problem also differs from Kasy (2012) which provided confidence sets for the number of solutions to the equation $g(y) = 0$, assuming that the function $g$ can be identified from some moment restrictions. To the best of my knowledge, Echenique and Komunjer (2009) is the only paper that considered nonparametric models with multiple equilibria and a continuous dependent variable. Instead of testing for multiple equilibria, they developed a test for complementarity between the dependent variable (at extremal values) and covariates. As a matter of dimension, Echenique and Komunjer (2009) focused on one dimensional equilibrium, (i.e., the dependent variable is a scalar). When the dependent variables in an economic model are multidimensional, (this happens in many empirical applications), they reduced the number of equilibrium characterizing equations by substitution. The difficulty with substitution is that it usually breaks the separable form of unobservables and observables in the resulting structural equation.[6] To avoid this problem, Echenique

[6] For example, if an equilibrium $(y_1, y_2)$ is characterized by $r_1(y_1, y_2) = u_1$ and $r_2(y_1, y_2) = u_2$, suppose the second equation yields the best response function of $y_2$ as $y_2 = \phi_2(y_1, u_2)$. By substitution, we get an equation $r_1(y_1, \phi_2(y_1, u_2)) = u_1$, where the unobservable $u_2$ does not appear as a separable form.

and Komunjer (2009) first assumed no unobservables in the structural equations, used substitution to reduce the dimension of equations, and then added a disturbance term back to the one-dimensional structural equation.[7] In this chapter, we consider the multidimensional dependent variables directly and keep unobservables in the original structural equations.

The form of the econometric model in this chapter resembles nonparametric simultaneous equation models (or the transformation model, when the dependent variable is a scalar) studied by Horowitz (1996), Ekeland, Heckman and Nesheim (2004), Matzkin (2008) and Berry and Halie (2013), among others. The crucial difference is that we do not impose the assumption that the unknown function $r(\cdot)$ is one to one. In this sense, this chapter proposes a test for one of the key assumptions maintained in the standard nonparametric simultaneous equations models.

Our test is also related to statistics literature on testing the presence of a jump in a density or a regression function. Such a test differs from the better known approach to detect the location of the jump given its existence. (For the latter, see Hall and Titterington (1992), Müller (1992) and Delgado and Hidalgo (2000), among others.) Chu and Cheng (1996) proposed a test for the presence of a jump in univariate densities. Their critical values were computed from a Gumbel distribution under the null hypothesis of continuity. A similar test for a univariate regression function has been developed by Hamrouni (1999). Müller and Stadtmüller (1999) showed that in univariate equidistant regression models, the sum of squared jump sizes can be represented as the coefficient of an asymptotic linear model, with the squared difference of $y$ as the dependent variable and twice the standard deviation of the error term as the intercept. Hence they developed a test by checking whether the sum of the squared jump sizes equals zero. Gijbels and Goderniaux (2004) proposed a two-step bootstrap test for discontinuities in univariate regression functions, which identified a discontinuity as a point with the largest derivative. Bowman, Pope and Ismail (2006) developed a test based on the sum of squared pointwise jumps for univariate and bivariate regression functions. Our test is based on maximized local jumps, which resembles Chu and Cheng (1996) and Qiu (2002). To compute the critical value, instead of using Gumbel distribution as an asymptotic approximation, we apply the

[7] For the aforementioned example, this means that they began with $r_1(y_1, y_2) = 0$ and $r_2(y_1, y_2) = 0$ in the first place, reduced it to $r_1(y_1, \phi_2(y_1)) = 0$, and then added a disturbance term $u$ to yield $r_1(y_1, \phi_2(y_1)) = u$.

Gaussian multiplier bootstrap proposed by Chernozhukov, Chetverikov and Kato (2013).

### 2.1.2 Organization of Chapter 2

This chapter is organized as follows. Section 2.2 presents the basic setup and provides examples. Section 2.3 establishes a link between uniqueness or multiplicity of equilibrium and the continuity or discontinuity in the density function of the dependent variables. Section 2.4 constructs a test and derives its asymptotic properties. Section 2.5 conducts Monte Carlo simulations. Section 2.6 concludes. All technical conditions and proofs are presented in Appendix A.

## 2.2 The Framework and Examples

### 2.2.1 The Econometric Model

Consider the following system of structural equations

$$r(Y) = g(X) + U, \tag{2.1}$$

where the dependent (endogenous) vector $Y$ has dimension $J$ and support $\mathcal{Y} \subset \mathbf{R}^J$. The vector-valued function $r(\cdot) : \mathcal{Y} \to \mathbf{R}^J$ is twice continuously differentiable. The exogenous vector $X$ has dimension $L$ and support $\mathcal{X} \subset \mathbf{R}^L$. The vector $g(X) = [g_1(X), g_2(X), ..., g_J(X)]'$, where the real-valued function $g_j : \mathcal{X} \to \mathbf{R}^J$ is twice continuously differentiable for each $j$. The functions $r$ and $g$ are nonparametric. In addition, we denote the support of $U$ as $\mathcal{U}$ and assume it to be connected. A generalization of model (2.1) is

$$r(Y, W) = g(X, W) + U, \tag{2.2}$$

where $(W, X)$ are observable exogenous variables. In the following, we focus on (2.1), since (2.2) can be reduced to (2.1) once we condition on $W$. The essential difference between (2.1) and the standard nonparametric simultaneous equations model is that we do not impose the assumption that function $r : \mathcal{Y} \to \mathbf{R}^J$ is one to one. This allows for the possibility of multiple $y$ satisfying (2.1) given exogenous $(x, u)$. When such multiplicity happens for some $(x, u)$, model (2.1) is incomplete. One can introduce a probability measure to account for the randomness of $Y$ given $(X, U)$. For that purpose, we first introduce some notations and impose Assumption D1. Let

$$\mathcal{E}(v) \equiv \left\{ y \in \mathbf{R}^J : r(y) = v \right\}, \text{ for any } v \in \mathcal{V},$$

where $\mathcal{V} \equiv \{g(x) + u : (x, u) \in \mathcal{X} \times \mathcal{U}\}$. $\mathcal{E}(v)$ is the set of solutions to the equation $r(y) = v$. In this chapter, we assume that $\mathcal{E}(v)$ is non-empty and has a finite cardinality for any $v \in \mathcal{V}$. This is Assumption D1.

**Assumption D1.** (i) $\mathcal{V} \subset \{v \in \mathbf{R}^J : r(y) = v, \text{for some } y \in \mathbf{R}^J\}$.
(ii) $\Pr(\#(\mathcal{E}(g(X) + U)) < \infty) = 1$, where $\#(G)$ is the cardinality of a generic set $G$.

If the range of $r$ function is $\mathbf{R}^J$, Assumption D1(i) always holds. Assumption D1(ii) rules out functions $r$ having a constant part such that $\#(\mathcal{E}(v))$ is uncountable for some $v$.

Now we describe the aforementioned probability that accounts for the randomness of $Y$ given $(X, U)$. For any $(x, u) \in \mathcal{X} \times \mathcal{U}$, let $\pi(x, u)$ be a probability on the power set of $\mathcal{E}(g(x) + u)$ specified by

$$\Pr(Y \in A | X = x, U = u) = \pi(x, u)(A), \tag{2.3}$$

for all subsets $A$ of $\mathcal{E}(g(x) + u)$.

Under Assumption D1, we classify the model (i.e., the structural equations (2.1) and the probability $\pi(x, u)$) into one of the following three categories based on equilibrium behaviour.

**(C1)** $\#(\mathcal{E}(g(x) + u)) = 1$ for all $(x, u) \in \mathcal{X} \times \mathcal{U}$.

**(C2)** $\Pr(\#(\mathcal{E}(g(X) + U)) > 1) > 0$ and $\Pr(\#(\text{supp}(\pi(X, U))) = 1) = 1$, where $\text{supp}(\pi(x, u))$ denotes the support of the probability $\pi(x, u)$.

**(C3)** $\Pr(\#(\mathcal{E}(g(X) + U)) > 1) > 0$ and $\Pr(\#(\text{supp}(\pi(X, U))) > 1) > 0$.

Category (C1) means that structural equations (2.1) admit a unique solution in $y$ for all $(x, u) \in \mathcal{X} \times \mathcal{U}$. Category (C2) means that structural equations (2.1) have more than one solutions in $y$ for $(x, u)$ with a positive probability, however, given $(x, u)$ there is only one element in the equilibria set $\mathcal{E}(g(x) + u)$ that $Y$ can take with a positive probability, and that probability is 1. Category (C3) means that structural equations (2.1) have more than one solutions in $y$ for some $(x, u)$ with a positive probability, and at those $(x, u)$ there are strictly more than one elements in $\mathcal{E}(g(x) + u)$ that $Y$ can take with positive probabilities. Throughout this chapter, a unique equilibrium refers to (C1) or (C2), whereas multiplicity of equilibria refers to (C3). In addition, we do not consider an exceptional case in which structural equations (2.1) admit multiple solutions for some $(x, u) \in \mathcal{X} \times \mathcal{U}$, but $\Pr(\#(\mathcal{E}(g(X) + U)) > 1) = 0$.

### 2.2.2 Examples

We present three examples:

**[Example 1] A Nonparametric BLP Model.** The model is based on Berry and Haile (2014). There are $T$ markets, $J$ firms and each firm produces a product. For the $t$th market, $p_t = (p_{1t}, ..., p_{Jt})'$ denotes product prices, $x_t = (x_{1t}, ..., x_{Jt})'$ denotes covariates in consumers' preferences, $w_t = (w_{1t}, ..., w_{Jt})'$ denotes covariates in cost functions, $s_t = (s_{1t}, ..., s_{Jt})'$ denotes market shares, $\zeta_t = (\zeta_{1t}, ..., \zeta_{Jt})'$ and $\omega_t = (\omega_{1t}, ..., \omega_{Jt})'$ denote product level unobservables in the preference and cost functions respectively. Conditional on $(x_t, \zeta_t)$, the utility for consumer $i$ $v_{it} = (v_{it0}, ...v_{it1}, ..., v_{itJ})'$ from consuming each product, (where $v_{i0t}$ denotes the utility of consumer $i$ from consuming nothing), is i.i.d. across consumers $i = 1, ..., I$ and markets $t = 1, ..., T$. The market share of product $j$ in market $t$ is the choice probability

$$s_{jt} = \sigma_j(p_t, \zeta_t, x_t) = \Pr\left(\arg\max_{k=0,1,...,J} v_{itk} = j\right).$$

According to Berry and Haile (2014), under assumptions including demand index restrictions[8] and connected substitutes[9], the nonparametric BLP model can be characterized by a system of $2J$ equations.

$$\sigma_j^{-1}(s_t, p_t) = x_{jt} + \zeta_{jt} \tag{2.4}$$

$$\pi_j^{-1}(s_t, p_t) = w_{jt} + \omega_{jt}, \tag{2.5}$$

for $j = 1, 2, .., J$. This system of equations corresponds to (2.1). Berry and Haile (2014) show that $\sigma_j^{-1}(\cdot)$ and $\pi_j^{-1}(\cdot)$ can be identified under "a unique equilibrium" assumption (Assumption 13 in Berry and Haile (2014)): "There is a unique vector of equilibrium prices associated with any $(\delta_{jt}, \kappa_{jt})$, (where $\delta_{jt} = x_{jt} + \zeta_{jt}$ and $\kappa_{jt} = w_{jt} + \omega_{jt}$)."[10] While imposing

[8] "Demand index" is Assumption 1 of Berry and Haile (2014), which leads to $s_{jt} = \sigma_j(p_t, \delta_t, x_t^{(2)})$, where $\delta_{jt} = x_{jt}^{(1)} + \zeta_{jt}$, $x_{jt}^{(1)}$ is one element of $x_{jt}$ and $x_{jt}^{(2)}$ are the remaining elements. As in their paper, $x_{jt}^{(2)}$ is suppressed because we can conditional on an arbitrary value of $x_{jt}^{(2)}$. For notational simplicity, write $\delta_{jt} = x_{jt} + \zeta_{jt}$.

[9] Connected substitutes (Assumption 2 of Berry and Haile (2014)) leads to $\delta_{jt} = \sigma_j^{-1}(p_t, s_t)$ for all $j = 1, ..., J$ and for any $(p_t, s_t)$.

[10] The model with "a unique equilibrium" in Berry and Haile (2014) corresponds to the union of Category (C1) and a subset of Category (C2). We include in the unique equilibrium (more precisely, category (C2)) the case where the equilibrium selection rule depends on $(x_t, \zeta_t)$.(Similarly, (C2) allows equilibrium selection rule depends on $(w_t, \omega_t)$.)

that assumption, they noted that "it is not hard to construct examples admitting multiple equilibria" and the unique equilibrium assumption "rules out random equilibrium selection or equilibrium selection based on $x_{jt}$ or $\zeta_{jt}$ instead of their sum $\delta_{jt}$ (and similarly for $\kappa_{jt}$)."

**[Example 2] A Semiparametric Social Interaction Model.** This is an extension of the standard linear-in-means model in social interaction (Manski (1993), Brock and Durlauf (2001b)). Here, the endogenous effect is captured by $\varphi(m_g)$, where the function $\varphi(\cdot)$ is not parametrically specified. The behavioural equation is

$$w_{gi} = c' x_{gi} + d' \bar{x}_g + e' y_g + \varphi(m_g) + \alpha_g + u_{gi}, \tag{2.6}$$

where $w_{gi}$ denotes individual outcome, $x_{gi}$ individual characteristics, $y_g$ group level characteristics, $\bar{x}_g = \mathbf{E}(x_{gi}|Y_g, \alpha_g, u_g)$ group level expected individual characteristics, $m_g = \mathbf{E}(w_{gi}|y_g, \alpha_g, u_g)$ expected group outcome, $\varphi(\cdot)$ the effect of expected group outcome on individual outcome, $\alpha_g$ unobservable group characteristics and $u_{gi}$ unobservable individual characteristics. The exogenous assumption is

$$\mathbf{E}\left(u_{gi}|x_{gi}, y_g, \alpha_g\right) = 0. \tag{2.7}$$

By a self-consistency condition, we have

$$m_g - \varphi(m_g) = \left(c' + d'\right) \bar{x}_g + e' Y_g + \alpha_g. \tag{2.8}$$

Equation (2.8) corresponds to (2.1). If we impose an ex-ante assumption that $\varphi^{'}(m) < 1$, there is a unique solution in $m_g$ given the observable and unobservable. Thus $\varphi(\cdot)$ is identified by the standard transformation model. Otherwise, there may be multiple equilibria in $m_g$.

**[Example 3] A macro model of employment**

Dagsvik and Javanovic (1994) set up a simple macro model which allows for multiple equilibria in employment rate. The structural equation characterizing an equilibrium employment rate $y$ is:

$$z - \beta \log \phi \left( \frac{1}{1 + \exp(-z)} \right) = \beta\delta \log x + \beta \log u - \log v,$$

The unique equilibrium assumption in in Berry and Haile (2014) corresponds to our Assumption D10 in Section 2.3. Thus by Proposition 3 of this chapter, we can use the continuity of the unconditional density to test the validity of the unique equilibrium assumption in Berry and Haile (2014).

where $z = \log(y/(1-y))$, $x$ denotes monetary supply, $u$ denotes a shock to aggregate demand or productivity, $v$ denotes aggregate shock to the supply of labour. Dagsvik and Javanovic (1994) assumed a cubic form of $\phi(\cdot)$ and a parametric form of equilibrium selection rule. In our framework, $\phi(\cdot)$ is not parametrically specified.

## 2.3 Testing the Presence of Multiple Equilibria via Discontinuity

This section analyses the relations between the uniqueness or multiplicity of equilibria and the continuity or discontinuity of the density function $f_{Y|X}(y|x)$ in $y$. In particular, under regularity conditions, a model with a unique equilibrium ((C1) or (C2)) necessarily makes the conditional density $f_{Y|X}(y|x)$ continuous over $y \in \text{int}(\mathcal{Y}(x))$, for all $x \in \mathcal{X}$, where $(\mathcal{Y}(x))$ is the support of $Y$ conditional on $X = x$. On the other hand, a model with multiple equilibria ((C3)), under reasonable conditions, gives rise to a discontinuity of $f_{Y|X}(y|x)$ in $y$. Here a discontinuity refers to a jump of $f_{Y|X}(y|x)$ between two *non-zero* values as we perturb $y$. Under an additional condition, a model with a unique equilibrium leads to continuous unconditional density $f_Y(y)$, while a model with multiple equilibria gives rise to a jump in the unconditional density $f_Y(y)$. Proofs in this section are collected in Appendix A.1.

### 2.3.1 An Equilibrium Selection Rule

The possibility of $\#(\mathcal{E}(g(x)+u)) > 1$ makes the model given by structural equations (2.1) incomplete in the sense that the realization of exogenous variables $(x, u)$ cannot uniquely determine the value of endogenous variables. We address this issue by introducing an equilibrium selection rule, which specifies the probability with which each "type" of equilibrium is selected[11]. Let us first give an index system to indicate types of equilibria. Towards this end, we begin with some notations and assumptions. Let

$$\bar{\mathcal{Y}} \equiv \{y \in \mathbf{R}^J : r(y) \in \mathcal{V}\}.$$

[11]It is possible to use $\pi(x, u)$ in (2.3) as an equilibrium selection rule, see Echenique and Komunjer (2009). However, we try a different approach here, which is more convenient for our analysis. Our equilibrium selection rule will be defined on "types" (or indices) of equilibria.

The set $\bar{\mathcal{Y}}$ collects all possible values of equilibria, and it is a superset of $\mathcal{Y}$ since some equilibria may not be realized. We impose a smoothness condition on functions $r$ and $g$.

**Assumption D2.** Functions $r : \bar{\mathcal{Y}} \to \mathbf{R}^J$ and $g : \mathcal{X} \to \mathbf{R}^J$ are twice continuously differentiable over $\mathrm{int}(\bar{\mathcal{Y}})$. The Jacobian of $r(y)$, denoted as $J(y)$, is bounded and twice continuously differentiable over $\mathrm{int}(\bar{\mathcal{Y}})$.

A large class of functions with an enough degree of smoothness satisfies Assumption D2.

If the structural equations (2.1) are in (C1), there is a unique reduced-form function which maps the exogenous term $v = g(x) + u$ to $y$. Otherwise, the relation from $v$ to $y$ turns to be a correspondence. Each branch of the correspondence can be viewed as a reduced-form function from $v$ to $y$. Definition 1 below formalizes it. The family $\{q_1(\cdot), ..., q_M(\cdot)\}$ can be viewed as a collection of reduced-from functions mapping an exogenous term $g(x)+u$ to a value of endogenous $Y$.

**Definition 1.** *Let $M$ be the smallest number such that there exists a family of twice continuously differentiable functions*

$$\{q_1(\cdot), q_2(\cdot), ..., q_M(\cdot)\},$$

*where $q_m : B_m \to \mathbf{R}^J$, $B_m$ is an open and connected subset of $\mathcal{V}$, and the following conditions are satisfied*:

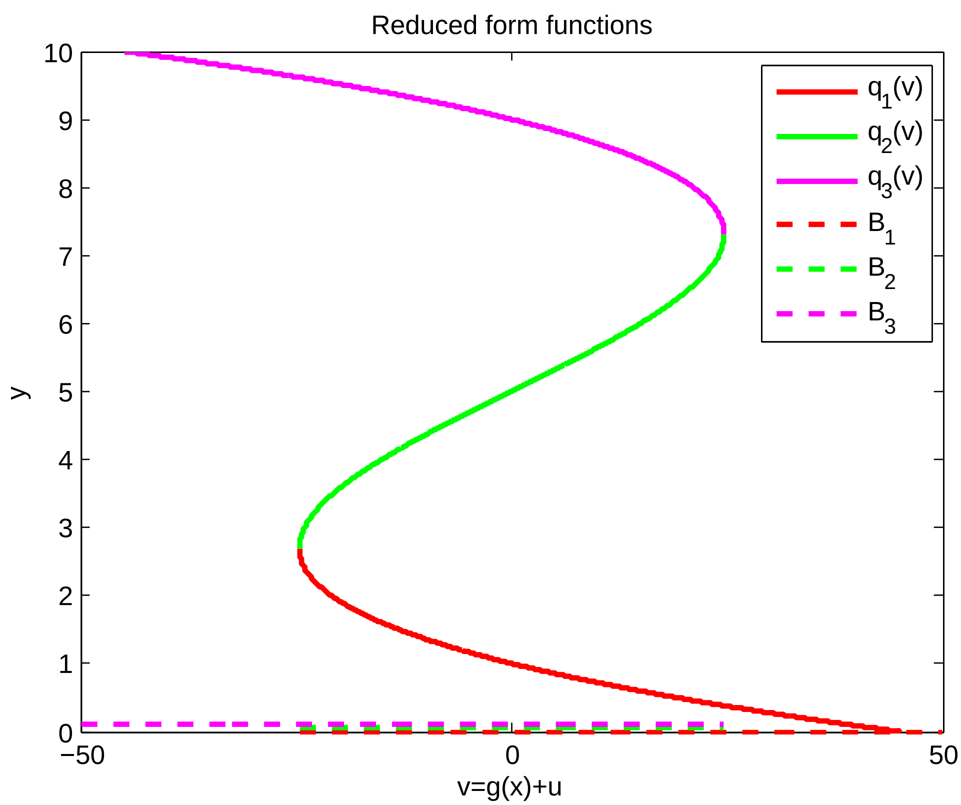
*(i) For any $(y, v) \in \bar{\mathcal{Y}} \times B_m$ satisfying $r(y) = v$, we have $y = q_m(v)$. Furthermore, $r(q_m(v)) = v$ for all $v \in B_m$.*

*(ii) $\cup_{m=1}^{M} \bar{B}_m = \mathcal{V}$. (Here $\bar{B}_m$ denotes the closure of $B_m$.)*

*(iii) For any $v \in B_m \cap B_k$ and $m \neq k$, $q_m(v) \neq q_k(v)$.*

Definition 1(i) implies that $q_m : B_m \to A_m$ is one to one in each $B_m$. (Suppose there are $v_1 \neq v_2$ in $B_m$ satisfying $y = q_m(v_1) = q_m(v_2)$. Thus we have $r(y) = v_1 \neq v_2 = r(y)$, which cannot hold.) Definition 1(ii) means that for almost all $v \in \mathcal{V}$, there is a function $q_m$ ($m$ may vary with $v$) defined on it. Definition 1(iii) says that each $q_m$ is distinct, otherwise we can exclude such a $v$ from either $B_m$ or $B_k$. The sets $B_1, ..., B_M$ are not all disjoint if the structural equations (2.1) admit multiple solutions at some $(x, u)$. Figure 2.2 illustrates the collection of reduced-form functions $\{q_1(\cdot), ..., q_M(\cdot)\}$ using a simple example. That is, the structural equation is $r(y) = u$, where

Figure 2.2: Reduced-form functions



$r(y) = -y^3 + 15y^2 - 59y + 45$. By Definition 1, the example has three reduced-form functions $q_1, q_2, q_3$ and their domains are $B_1, B_2$ and $B_3$.

We impose the following assumption on the structural equations (2.1).

**Assumption D3.** (i) The constant $M$ in Definition 1 is finite. (ii) If structural equations (2.1) admit a unique solution in $y$ for all $(x, u) \in \mathcal{X} \times \mathcal{U}$, the Jacobian $J(y) > 0$ for all $y \in \bar{\mathcal{Y}}$.

Assumption D3 (i) requires that the collection of reduced-form function mapping from $g(x) = u$ to $y$ has finite elements. Assumption D3 (ii) guarantees that for any model in (C1), the Jacobian $J(y)$ is bounded away from zero.

Let $A_m = q_m(B_m)$, the family of the reduced-form functions $\{q_1, q_2, ..., q_M\}$ in Definition 1 produces a collection of subsets of $\bar{\mathcal{Y}}$ with the properties in Lemma 1 below.

**Lemma 1.** *Let $A = \{A_m, m = 1, 2, ..., M\}$, where $A_m = q_m(B_m)$. Then under Assumptions D1-D3, the collection $A$ has the following properties:*

*(i) Function $r$ is one to one on each $A_m$.*

*(ii) $\bar{\mathcal{Y}} = \cup_{m=1}^{M} \bar{A}_m$.*

*(iii) For any $m \neq k$, $A_m \cap A_k = \varnothing$.*

*(iv) If a model is in (C1), the constant $M = 1$* in *Definition 1.*

*(v) If a model is in (C2) or (C3), the constant $M > 1$ in Definition 1.*

Lemma 1 is a direct consequence of Assumptions D1-D3 and Definition 1. The key role of Lemma 1 is to generate an index system of equilibria types. If $y \in A_m$, we call it the type $m$ equilibrium and index it by $m$. Define an equilibria indices set (given the exogenous value $v = g(x) + u$) as a collection of $m$'s such that (2.1) is satisfied for some $y \in A_m$. That is, for any fixed $v \in \mathcal{V}$,

$$\mathcal{M}(v) \equiv \{m : v \in r(A_m)\}. \tag{2.9}$$

Note that the function $\mathcal{M} : \mathcal{V} \rightarrow$ the power set of $\{1, 2, ..., M\}$ specified in (2.9) is fully determined by the structual equations (2.1). Let

$$\mathcal{W} = \{(x, u) : g(x) + u \in \cup_{m=1}^{M} B_m\},$$

we introduce the equilibrium selection rule.

**Definition 2.** *Let an equilibrium selection rule $\lambda$ be a measurable function $\mathcal{W} \rightarrow [0, 1]^M$ defined as*

$$\lambda(x,u) = [\lambda_1(x,u), ..., \lambda_M(x,u)],$$

*where*

$$\lambda_m(x,u) = \Pr\left(Y \in A_m | X = x, U = u\right).$$

Clearly, $\lambda_h(x,u) = 0$ if $h \notin \mathcal{M}(g(x)+u)$, and $\sum_{m=1}^{M} \lambda_m(x,u) = 1$. Given a function $\mathcal{M}$, the collection of all equilibrium selection rules are

$$\mathcal{C}(\mathcal{M}) = \left\{ \begin{array}{l} \lambda : \mathcal{W} \to [0,1]^M : \sum_{m=1}^{M} \lambda_m(x,u) = 1, \\ \lambda_h(x,u) = 0 \text{ if } h \notin \mathcal{M}(g(x)+u), \text{ for all } (x,u) \in \mathcal{W} \end{array} \right\}.$$

In some situations, the econometrician may impose restrictions on the equilibrium selection rule. For example, later we will consider a class $\mathcal{C}_0(\mathcal{M})$ of equilibrium selection rules that no longer depends on unobservables, conditional on the equilibria indices set and the observable characteristics. That is, let

$$\mathbf{M} = \{\mathcal{M}(v) : v \in \cup_{m=1}^{M} B_m\}, \tag{2.10}$$

we have

$$\mathcal{C}_0(\mathcal{M}) = \left\{ \begin{array}{l} \lambda \in \mathcal{C}(\mathcal{M}) : \text{there is a } \phi : \mathbf{M} \times \mathcal{X} \to [0,1]^M \\ \text{such that } \lambda(x,u) = \phi(\mathcal{M}(g(x)+u), x), \text{ for all } (x,u) \in \mathcal{W} \end{array} \right\}. \tag{2.11}$$

We say an equilibrium selection rule $\lambda$ is degenerate at $(x,u)$ if $\lambda_m(x,u) = 1$ for an $m$. Now by Lemma 1 and Definition 2, we obtain equivalent statements of (C1) to (C3) about the equilibrium behaviour of the model (the structural equations (2.1) and the equilibrium selection rule $\lambda$).

**(C1)** $M = 1$.
**(C2)** $M > 1$, and $\Pr\left(\lambda_m(X,U) = 1 \text{ for some } m\right) = 1$.
**(C3)** $M > 1$, and $\Pr\left(\lambda_m(X,U) < 1 \text{ for all } m\right) > 0$.

### 2.3.2 A Testing Criterion

We assume that the unobservable $U$ has a smooth density function.

**Assumption D4.** (i) The unobservable $U$ is independent of $X$. (ii) The distribution of $U$ is absolutely continuous with respect to the Lebesgue measure and its density $f_U(u) = dF_U(u)/du$ is bounded and twice continuously differentiable. (iii) The support $\mathcal{U}$ is connected.

Assumption D4 means that the unobservable vector $U$ is independent of the observable characteristics and its density is smooth enough. Therefore, any jump in the density of dependent variables is not due to the unobservable $U$. In the next subsection, we will relax this assumption to allow for jumps in $f_U(u)$.

When the structural equations (2.1) have more than one solutions, a typical situation is that the equilibria indices set $\mathcal{M}(v)$ defined in (2.9) changes with $v$. This is the main source for a jump in the conditional density functions. Suppose there are $y_1, y_2 \in A_m$ associated with distinct equilibria indices sets. (i.e., $\mathcal{M}(r(y_1)) \neq \mathcal{M}(r(y_2))$.) Based on that, we can further divide $A_m$ into disjoint subsets. To formalize this idea, recall that $\mathbf{M}$ in (2.10) is the collection of all equilibria indices sets generated by the structural equation (2.1). Let $\mathbf{M}_m$ be the subset of $\mathbf{M}$ such that each equilibria indices set in $\mathbf{M}_m$ equal to $\mathcal{M}(r(y))$ for some $y \in A_m$:

$$\mathbf{M}_m = \{G \in \mathbf{M} : G = \mathcal{M}(r(y)), \text{ for some } y \in A_m\}.$$

Let $L_m = \{1, ..., \#(\mathbf{M}_m)\}$ so that we can list the elements in $\mathbf{M}_m$:

$$\mathbf{M}_m = \{G_l^m : l \in L_m\}.$$

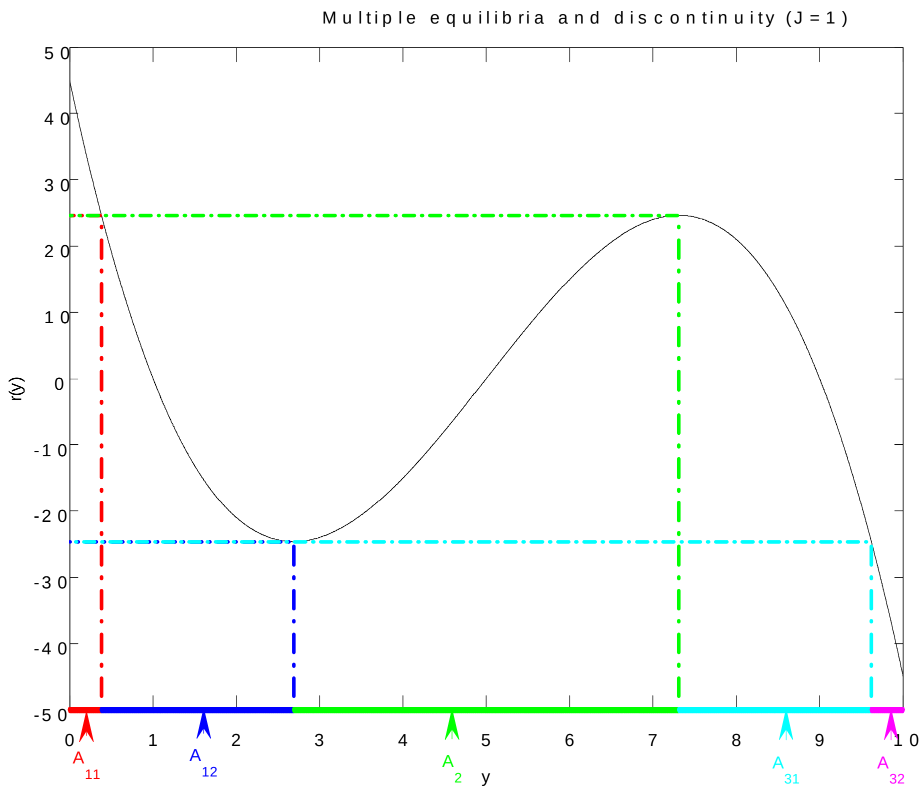
Using this notation, for each $m = 1, ..., M$, we can produce a collection of subsets of $A_m$ based on equilibria indices sets in $\mathbf{M}_m$,

$$A_{ml} = \{y \in A_m : \mathcal{M}(r(y)) = G_l^m\}, \text{ for } l \in L_m. \tag{2.12}$$

By construction, we have (i) $\bar{A}_m = \cup_{l \in L_m} \bar{A}_{ml}$, (ii) $A_{ml} \cap A_{ml'} = \varnothing$, (iii) $\mathcal{M}(r(A_{ml}))$ is defined, and (iv) $\mathcal{M}(r(A_{ml})) \neq \mathcal{M}(r(A_{ml'}))$ for all $m$ and $l \neq l'$. Note that (iii) and (iv) uses the following notation: for any $V \subset \mathcal{V}$, define $\mathcal{M}(V) = \mathcal{M}(v)$ if $\mathcal{M}(v)$ is the same over $v \in V$.

Consider the previous simple example without covariates $X$: $r(y) = u$, where $r(y) = -y^3 + 15y^2 - 59y + 45$. In Figure 2.3, $r(y)$ is plotted against

Figure 2.3: Multiple equilibria and jump(s)



$y$. There can be one or three equilibria (with positive probabilities). In this example, the three reduced-form functions illustrated in Figure 2.2 divides $\bar{\mathcal{Y}}$ into $A_1, A_2$ and $A_3$. Each of them is labelled with a different colour on the horizontal axis. Function $r(y)$ is one to one within each $A_m$. An equilibrium is indexed as type $m$ if it is within $A_m$, $m = 1, 2, 3$. Furthermore, we have $\mathbf{M} = \{\{1\}, \{3\}, \{1, 2, 3\}\}$, and $\mathbf{M}_1 = \{\{1\}, \{1, 2, 3\}\}$, $\mathbf{M}_2 = \{\{1, 2, 3\}\}$, $\mathbf{M}_3 = \{\{3\}, \{1, 2, 3\}\}$. Based on distinct equilibria indices sets (corresponding to (2.12)), we have $\bar{A}_1 = \bar{A}_{11} \cup \bar{A}_{12}$ and $\bar{A}_3 = \bar{A}_{31} \cup \bar{A}_{32}$, where the equilibrium indices sets associated with $A_{11}, A_{12}, A_2, A_{31}$ and $A_{32}$ are:

$$\mathcal{M}(r(A_{11})) = \{1\}, \mathcal{M}(r(A_{32})) = \{3\},$$
$$\mathcal{M}(r(A_{12})) = \mathcal{M}(r(A_2)) = \mathcal{M}(r(A_{31})) = \{1, 2, 3\}.$$

Now we begin to analyse the conditional density function $f_{Y|X}(y|x)$ and its relation to the uniqueness/multiplicity of equilibria. Lemma 2 gives a formula of the conditional density function.

**Lemma 2.** *Suppose that Assumptions D1-D4 hold. Then for any $y \in \cup_{m=1}^{M} \cup_{l \in L_m} A_{ml}$ and $x \in \mathcal{X}$,*

$$f_{Y|X}(y|x) = f_U(r(y) - g(x))|J(y)|\lambda_s(x, r(y) - g(x)),$$

*where $s \in \{1, 2, ..., M\}$ and $h \in L_s$ such that $y \in A_{sh}$.*

Lemma 2 shows that under some regularity conditions, the conditional density $f_{Y|X}(y|x)$ can be written as a product of the density of unobserved $U$, the Jacobian of $r(y)$, and a selection probability (a component of the equilibrium selection rule). Since $A_{ml}$'s are mutually exclusive across $m$ and $l$, for any $y \in \cup_{m=1}^{M} \cup_{l \in L_m} A_{ml}$, there is a unique pair $(s, h)$ where $s \in \{1, 2, ..., M\}$ and $h \in L_m$, such that $y \in A_{sh}$. Also, note that $\cup_{m=1}^{M} \cup_{l \in L_m} \bar{A}_{ml} = \bar{\mathcal{Y}}$. From Assumption D2, D4 and Lemma 2, we can see that a jump in $f_{Y|X}(y|x)$, if it exists, must come from the selection probability specified by the equilibrium selection rule. The following proposition establishes a testable implication of a unique equilibrium.

**Proposition 1.** *Suppose Assumptions D1-D4 hold, and the model (i.e., the structural equations (2.1) and the equilibrium selection rule $\lambda$) generates a unique equilibrium (i.e., the model falls into Category (C1) or (C2)). Then*

*(i) The conditional density $f_{Y|X}(y|x)$ is continuous in $y$ on $int(\mathcal{Y}(x))$ given any $x \in \mathcal{X}$. ($\mathcal{Y}(x)$ is the support of $\mathcal{Y}$ given $x$.) That is, for any $x \in \mathcal{X}$,*

*$y \in int(\mathcal{Y}(x))$ and any two sequences $\{y_{1n}\}$, $\{y_{2n}\}$ such that $y_{1n}, y_{2n} \in \mathcal{Y}(x), y_{1n} \to y$ and $y_{2n} \to y$, we have*

$$\lim_{y_{1n} \to y} f_{Y|X}(y_{1n}|x) = \lim_{y_{2n} \to y} f_{Y|X}(y_{2n}|x).$$

*(ii) If $X$ is a continuous random vector, $f_{Y|X}(y|x)$ is continuous in $(x, y)$ over $x \in int(\mathcal{X})$, $y \in int(\mathcal{Y}(x))$. That is, for any $x \in int(\mathcal{X})$, $y \in int(\mathcal{Y}(x))$ and any two sequences $\{(y_{1n}, x_{1n})\}$, $\{(y_{2n}, x_{2n})\}$ such that $y_{1n} \in \mathcal{Y}(x_{1n}), y_{2n} \in \mathcal{Y}(x_{2n}), (y_{1n}, x_{1n}) \to (y, x)$ and $(y_{2n}, x_{2n}) \to (y, x)$, we have*

$$\lim_{(y_{1n}, x_{1n}) \to (y,x)} f_{Y|X}(y_{1n}|x_{1n}) = \lim_{(y_{2n}, x_{2n}) \to (y,x)} f_{Y|X}(y_{2n}|x_{2n}).$$

**Remark 1.** (i) Here continuity at $y$ means that the limits to $y$ from all directions in the support coincide. It does not impose any restriction on the value of the density function at $y$. Throughout this chapter, "continuity" refers to "lack of a jump discontinuity".

(ii) Conditional density $f_{Y|X}(y|x)$ may have jumps in $y$ from some positive number to zero at the boundary of $\mathcal{Y}(x)$. Therefore Proposition 1 is stated for $y \in int(\mathcal{Y}(x))$. Later, when constructing the test statistic, we exclude data points near the boundary of its support.

(iii) Though the main testing criterion of this chapter is a jump of the density function in $y$, Proposition 1(ii) tells us that a jump in $x$ actually provides additional evidence for multiple equilibria.

(iv) If the model is in (C1), it will produce another restriction on the observable data $(Y, X)$. That is, the support $\mathcal{Y}(x)$ is connected for all $x$. In particular, when the model is in (C1), function $r$ is one to one, hence $Y = r^{-1}(g(X) + U)$. Note that $\mathcal{U}$ is connected, $U$ is independent of $X$ (Assumption D4) and $r$ is continuous (Assumption D2). Therefore, for any $x$, $\mathcal{Y}(x)$ is connected by Theorem 23.5 of Munkres (1999). (That is, the image of a connected space under a continuous map is connected.) For example, if assuming $\mathcal{U} = \mathbf{R}^J$, we can further distinguish a model in (C1) from another model in (C2), because (C1) leads to $\mathcal{Y} = \mathbf{R}^J$ whereas (C2) produces hole(s) in $\mathcal{Y}$.

Proposition 1 states that a model with a unique equilibrium necessarily leads to continuous conditional density $f_{Y|X}(y|x)$. The next questions are whether and when a model with multiple equilibria would produce a jump of $f_{Y|X}(y|x)$ in $y$. The intuition is that if a model has multiple equilibria,

Table 2.1: Conditions for multiple equilibria to produce jump(s) in density

| Results | Assumptions on | |
|---|---|---|
| | Structural equations (2.1) | Equilibrium selection rule |
| Proposition 2 | D5(i), D6(i) | D6(ii) |
| Corollary 2 | D5, D7(i) | D7(ii), D8 |
| Corollary 3 | D5, D7(i), D9(i) | D7(ii), D9(ii) |

one would typically see that the equilibria indices set $\mathcal{M}(r(y))$ changes with $y$. Then at a fixed $x$, consider what would happen at the boundary between regions of $y$ associated with different equilibria indices sets. Suppose there is a region with a unique equilibrium and another region with multiple equilibria. At one side of the boundary, the selection probability is one, while on the other side of the boundary, the selection probability tends to be below one because the equilibrium selection rule has to pick one from multiple candidates. As a result, the selection probability tends to have a jump at the boundary between subsets of $\bar{\mathcal{Y}}$ with distinct equilibria indices sets. This leads to a jump of $f_{Y|X}(y|x)$ in $y$.

In the following, each of Proposition 2, Corollary 2 and Corollary 3 establishes the relation between multiple equilibria and jump(s) of $f_{Y|X}(y|x)$ in $y$, under different sets of assumptions. As an easier reference, Table 2.1 lists three sets of assumptions sufficient for multiple equilibria to produce such jump(s) in the conditional density. Keep in mind that Assumptions D1-D4 are maintained for all three results in Table 2.1.

The following Assumption D5 is a technical assumption. Assumption D5(i) will be used in Proposition 2, while Assumption D5(ii) and (iii) will be used later to give a sufficient condition for the equilibria indices set to vary on some $r(A_m)$.

**Assumption D5.** (i) For any $\varepsilon > 0$, there exists a $\varsigma > 0$ such that $J(y) > \varsigma$ for all $y \in \mathring{A}_m \subset A_m$ and all $m$, where $\mathring{A}_m$ satisfies $\mu_L(A_m \backslash \mathring{A}_m) < \varepsilon$ and $\mu_L$ denotes the Lebesgue measure on $\bar{\mathcal{Y}}$.

(ii) For any $v \in \operatorname{int}(\mathcal{V})$, if for any $\epsilon > 0$ there exists a $v_0$ such that $|v - v_0| < \epsilon$ and $r(y) = v_0$ has a unique solution in $y$, then the Jacobian $J(y) > 0$ for all $y$ satisfying $r(y) = v$.

(iii) For any $m \in \{1, ..., M\}$, let $v_0$ be an arbitrary point $v_0 \in \bar{B}_m \backslash B_m$, and $y^*$ satisfy $r(y^*) = v_0$, $y^* \neq \lim_{v \to v_0, v \in B_m} q_m(v) \equiv y_m$. If $y^*$ exists and there is no such $y'$ that every component of $y' - y_m$ has the same sign as $y^* - y_m$ and $|y' - y_m| < |y^* - y_m|$, then $J(y^*) > 0$.

Assumption D5(i) means that the Jacobian of function $r$ is bounded away from zero over a large enough subset of $A_m$, for all $m$. Assumption D5(ii) and (iii) bound the Jacobian $J(y)$ away from zero at some particular points.

Suppose Assumptions D1-D4 and D5(i) hold, the following Assumption D6 is sufficient for multiple equilibria to produce jump(s) in the conditional density of dependent variables.

**Assumption D6**. If the model generates multiple equilibria (i.e., the model falls in Category (C3)), there exists an $s \in \{1, 2, ..., M\}$ satisfying the following conditions:

(i) The set $A_s$ has disjoint subsets $A_{sh}$ and $A_{sh'}$ satisfying $\bar{A}_{sh} \cap \bar{A}_{sh'} \neq \varnothing$ and $\mathcal{M}(r(A_{sh})) \neq \mathcal{M}(r(A_{sh'}))$.

(ii) For some $y \in \bar{A}_{sh} \cap \bar{A}_{sh'}$, where $A_{sh}$ and $A_{sh'}$ satisfy Assumption D6(i), there is a set $\mathcal{X}_D \subset \mathcal{X}$ with $\Pr(X \in \mathcal{X}_D) > 0$ such that the following conditions hold for any $x \in \mathcal{X}_D$,

$$\lim_{y' \in A_{sh}, y' \to y} \lambda_s \left(x, r(y') - g(x)\right) \neq \lim_{y' \in A_{sh'}, y' \to y} \lambda_s \left(x, r(y') - g(x)\right),$$

and

$$\lim_{y' \in A_{sh}, y' \to y} \lambda_s \left(x, r(y') - g(x)\right) \cdot \lim_{y' \in A_{sh'}, y' \to y} \lambda_s \left(x, r(y') - g(x)\right) \neq 0.$$

**Remark 2.** (i) If the equilibrium selection rule belongs to $\mathcal{C}_0(\mathcal{M})$ in (2.11), (That is, it does not depend on $u$, given $\mathcal{M}(g(x) + u)$ and $x$.) Assumption D6(ii) becomes: for some $y \in \bar{A}_{sh} \cap \bar{A}_{sh'}$, where $A_{sh}$ and $A_{sh'}$ satisfy Assumption D6 (i), there is a set $\mathcal{X}_D \subset \mathcal{X}$ with $\Pr(X \in \mathcal{X}_D) > 0$ such that the following conditions hold:

$$\phi_s\left(\mathcal{M}(r(A_{sh})), x\right) \neq \phi_s\left(\mathcal{M}(r(A_{sh'})), x\right),$$
$$\phi_s\left(\mathcal{M}(r(A_{sh})), x\right) \cdot \phi_s\left(\mathcal{M}(r(A_{sh'})), x\right) \neq 0,$$

for all $x \in \mathcal{X}_D$.

(ii) Assumption D6 outlines the direct source for a jump (between non-zero values) of $f_{Y|X}(y|x)$ as we perturb $y$. Note that $A_{sh}$ and $A_{sh'}$ are associated with distinct equilibria indices sets, $\mathcal{M}(r(A_{sh})) \neq \mathcal{M}(r(A_{sh'}))$. Assumption D6 assumes two properties for a model with multiple equilibria: (i) The equilibria indices set $\mathcal{M}(r(y))$ varies over $\bar{\mathcal{Y}}$. (ii) A component of the equilibrium selection rule has a jump (between non-zero values) at the boundary between some neighbouring subsets of $\bar{\mathcal{Y}}$ with distinct equilibria indices sets. Assumption D6 reasonably holds in many interesting cases with multiple equilibria. For example, Corollary 1 below states that if Assumption D5(ii), (iii) hold, and the structural equations admit a unique solution for some values of exogenous variables while admit multiple solutions for other values, Assumption D6(i) is necessarily satisfied. For our example in Figure 2, the equilibria indices set changes at the boundary between $A_{11}$ and $A_{12}$ as well as at the boundary between $A_{31}$ and $A_{32}$. Imposing Assumption D6(ii) in that example includes the case where the equilibrium selection rule assigns to equilibrium 1 a probability strictly between zero and one when the equilibria indices set contains equilibrium $1, 2$ and 3.

(iii) Suppose that Assumption 6(i) hold. Then the equilibrium selection rules that do not satisfy Assumption D6(ii) can only take values of a zero measure "around" $(x, r(y)-g(x))$ for all $x \in \mathcal{X}$ and $y \in \bar{A}_{sh} \cap \bar{A}_{sh'}$. Precisely, let the equilibrium selection rules that do not satisfy Assumption D6(ii) form a collection $\mathcal{C}_E(\mathcal{M})$. It must have the following property:

$$\left\{ \begin{pmatrix} \lim_{y' \in A_{sh}, y' \to y} \lambda_s\left(x, r(y') - g(x)\right) \\ \lim_{y' \in A_{sh'}, y' \to y} \lambda_s\left(x, r(y') - g(x)\right) \end{pmatrix} : \lambda \in \mathcal{C}_E(\mathcal{M}) \right\}$$

has a zero measure in

$$\left\{ \begin{pmatrix} \lim_{y' \in A_{sh}, y' \to y} \lambda_s\left(x, r(y') - g(x)\right) \\ \lim_{y' \in A_{sh'}, y' \to y} \lambda_s\left(x, r(y') - g(x)\right) \end{pmatrix} : \lambda \in \mathcal{C}(\mathcal{M}) \right\},$$

for all $x \in \mathcal{X}$, $y \in \bar{A}_{sh} \cap \bar{A}_{sh'}$ and all $s, h, h'$ satisfying $\bar{A}_{sh} \cap \bar{A}_{sh'} \neq \varnothing$ and $\mathcal{M}(r(A_{sh})) \neq \mathcal{M}(r(A_{sh'}))$. (See Assumption D6(i).)

The following Proposition 2 establishes a testable implication for models that have multiple equilibria and satisfy Assumption D6.

**Proposition 2.** *Suppose Assumptions D1-D4, D5(i) and D6 hold and the model generates multiple equilibria (i.e., the model falls into Category (C3)). Then for all* $x \in \mathcal{X}_D$,

*(i) The conditional density* $f_{Y|X}(y|x)$ *has a jump at some* $y \in int(\mathcal{Y}(x))$. *That is, there are disjoint open subsets* $C_1, C_2 \subset int(\mathcal{Y}(x))$ *with* $\bar{C}_1 \cap \bar{C}_2 \neq \varnothing$

*and* $y_d = \bar{C}_1 \cap \bar{C}_2$ *satisfying*

$$\left| \lim_{y \in C_1, y \to y_d} f_{Y|X}(y|x) - \lim_{y \in C_2, y \to y_d} f_{Y|X}(y|x) \right| = \delta > 0.$$

*(ii) If* $X$ *is a continuous random vector,* $f_{Y|X}(y|x)$ *has a jump at* $(x, y)$ *for some* $x \in int(\mathcal{X})$, $y \in int(\mathcal{Y}(x))$.

In the following, we consider a typical case of multiple equilibria (to be specified in Assumption D7). That is, (i) the structural equations (2.1) admit a unique solution for some values of the exogenous variables, and have multiple solutions for other values, and (ii) the equilibrium selection rule $\lambda \in \mathcal{C}_0(\mathcal{M})$ in (2.11)(random selection, for example). We state the relations from multiple equilibria to the presence of a jump of density $f_{Y|X}(y|x)$ in $y$ for models with such structural equations and equilibrium selection rules.

**Assumption D7.** If the model generates multiple equilibria, the following conditions are satisfied.
(i) $\Pr(g(X) + U \in S) > 0$, where $S = \{v \in \mathcal{V} : \#(\mathcal{M}(v)) = 1\}$.
(ii) The equilibrium selection rule $\lambda \in \mathcal{C}_0(\mathcal{M})$ as defined in (2.11).

Assumption D7(i) means that the structural equations admit a unique solution for some values of exogenous variables, while they generate multiple solutions for other values. Assumption D7(ii) means that the only way the unobservable $u$ affects the equilibrium selection rule is through the equilibrium index set $\mathcal{M}(g(x) + u)$. In other words, given $\mathcal{M}(g(x) + u)$, the selection rule is purely random or it only depends on the observable covariates $x$. Assumption D7(i) is satisfied in several models with multiple equilibria, see Dagsvik and Jovanovic (1994) and Echenique and Komunjer (2007). It is reasonable that when the exogenous term $g(x) + u$ takes a very large (or small) value, its effect dominates the self-fulling mechanism that may lead to multiple equilibria, so that the exogenous variables uniquely determine the dependent variables. In contrast, when the exogenous term $g(x) + u$ takes a moderate value, the self-fulling mechanism tends to dominate and may produce (for example) three equilibria: a large one, a medium one and a small one.

Assumption D7(ii) says that the equilibrium selection rule $\lambda \in \mathcal{C}_0(\mathcal{M})$ as defined in (2.11). Echenique and Komunjer (2009) made an assumption very similar to Assumption D7(ii) when introducing their equilibrium selection rule.[12] Under Assumption D7(ii), for all $(x,u) \in \mathcal{W}$, we have $\lambda(x,u) = \phi(\mathcal{M}(g(x)+u),x) = \phi(\mathcal{M}(r(y)),x) = \phi(\mathcal{M}(r(C)),x)$, where $y \in C$, $r(y) = g(x)+u$, and $\mathcal{M}(r(C))$ is defined. In the following, whenever Assumption 7(ii) holds, we will use $\phi_m(\mathcal{M}(r(C)),x)$ to denote the $m$th component of the equilibrium selection rule.

In addition, let $\mathcal{C}_{E0}(\mathcal{M})$ be the collection of equilibrium selection rules that satisfy Assumption D7(ii) but do not satisfy Assumption D6(ii) (i.e., may not produce a jump in the conditional density when there are multiple equilibria). By Assumption D6, D7(i) and Remark 2, we must have

$$\{\phi_s\left(\mathcal{M}(r(A_{sh})),x\right), \phi_s\left(\mathcal{M}(r(A_{sh'})),x\right)) : \phi \in \mathcal{C}_{E0}(\mathcal{M})\}$$

has a zero measure in

$$\{\phi_s\left(\mathcal{M}(r(A_{sh})),x\right), \phi_s\left(\mathcal{M}(r(A_{sh'})),x\right)) : \phi \in \mathcal{C}_{0}(\mathcal{M})\},$$

for all $x \in \mathcal{X}$, and all $s, h, h'$ such that $\bar{A}_{sh} \cap \bar{A}_{sh'} \neq \varnothing$ and $\mathcal{M}(r(A_{sh})) \neq \mathcal{M}(r(A_{sh'}))$. The existence of such $s, h, h'$ is ensured by Assumption D6(i), or by the following Corollary 1, which states that Assumption D7(i) is sufficient for Assumption D6(i), under technical conditions Assumption D5(ii) and (iii).

**Corollary 1.** *Suppose Assumptions D1, D2, D5(ii) and (iii) hold. If the model (2.1) generates multiple equilibria, Assumption D7(i) is sufficient for Assumption D6(i).*

Now we give sufficient conditions for Assumption D6(ii).

**Assumption D8.** Suppose the model generates multiple equilibria, and Assumption D7 holds. Then there exists an $m \in \{1,2,...,M\}$ satisfying $r(A_m) \cap S \neq \varnothing$ ($S$ is defined in Assumption D7), and an $\mathcal{X}_D \subset \mathcal{X}$, $\Pr(X \in \mathcal{X}_D) > 0$, such that the following two conditions hold.

(i) For all $C \subset A_m$ such that $\mathcal{M}(r(C))$ is defined, the $m$th component of the equilibrium selection rule satisfies

$$\phi_m(\mathcal{M}(r(C)),x) > 0.$$

(ii) For some $C_0 \subset A_m$ such that $\mathcal{M}(r(C_0))$ is defined, we have

$$\phi_m(\mathcal{M}(r(C_0)),x) < 1.$$

[12]They wrote: "For a given $x$, different realizations of $u$ can affect the support of $\mathcal{P}_{xu}$, but not the probabilities assigned to different outcomes in the support."

Assumption D8 assumes that there is an $A_m \subset \bar{\mathcal{Y}}$ such that a subset of $A_m$ is associated with the singleton equilibrium indices set $\{m\}$, while another subset of $A_m$ is associated with a larger equilibria indices set (which also contains the element $m$). Furthermore, within $A_m$, the probability of choosing equilibrium $m$ is always positive, and sometimes is strictly less than one. Corollary 2 gives another version of Proposition 2 when Assumption D6 is replaced by Assumption D7 and D8.

**Corollary 2.** *Suppose the model generates multiple equilibria, and Assumptions D1-D5, D7 and D8 hold. Then there is a set $\mathcal{X}_{D_1} \subset \mathcal{X}$ with* $\Pr(X \in \mathcal{X}_{D_1}) > 0$, *such that for all $x \in \mathcal{X}_{D_1}$ the conclusions of Proposition 2 hold.*

The example in Figure 2 satisfies Assumption D7(i). Recall that there is no $X$. Thus the structural equations are $r(y) = u$ and the equilibrium selection rule $\lambda$ is a function of $u$. As a result, it will satisfy Assumption D7(ii) if its equilibrium selection rule satisfies that

$$\lambda\left(r(y)\right) = \phi\left(\mathcal{M}(r(A_{mh}))\right),$$

for all $y \in A_{mh}$, all $m$ and $h \in L_m$. Furthermore, it is not hard to see that Assumption D8 also holds in that example. Indeed, the $m$'s satisfying $r(A_m) \cap S \neq \varnothing$ are $m = 1$ and 3. Suppose Assumption D8 is violated, that is, $\phi_1\left(\mathcal{M}(r(A_{12}))\right)$ and $\phi_3\left(\mathcal{M}(r(A_{31}))\right)$ are either 0 or 1. Then there are three cases, (recall $\phi_1\left(\mathcal{M}(r(A_{12}))\right) + \phi_3\left(\mathcal{M}(r(A_{31}))\right) \leq 1$, which means the two terms cannot all equal 1), however, all of them fall into Category (C2), which corresponds to a unique equilibrium. As a result, when there are multiple equilibria, at least one of $\phi_1\left(\mathcal{M}(r(A_{12}))\right)$ and $\phi_1\left(\mathcal{M}(r(A_{31}))\right)$ must be in the open interval $(0,1)$. Suppose $0 < \phi_1\left(\mathcal{M}(r(A_{12}))\right) < 1$, density $f_Y(y)$ will have a jump at $y_{d1} = \bar{A}_{11} \cap \bar{A}_{12}$ (i.e, the boundary between $A_{11}$ and $A_{12}$).

The following Assumption D9 spells out an alternative set of sufficient conditions for Assumption D6. This time we do not assume that a component in the equilibrium selection rule is strictly less than 1.

**Assumption D9.** Suppose that the model generates multiple equilibria and Assumption D7 holds. Then,

(i) For any $v \in \mathcal{V}$, there exists an $m \in \{1, 2, ..., M\}$ such that $v \in r(A_m)$ and $r(A_m) \cap S \neq \varnothing$. ($S$ is defined in Assumption D7.)

(ii) For any $m$ satisfying $r(A_m) \cap S \neq \varnothing$, we have $\phi_m(\mathcal{M}(r(C)), x) > 0$ for all $C \subset A_m$ with $\mathcal{M}(r(C))$ defined, and for all $x \in \mathcal{X}$.

Assumption D9(i) is imposed on the structural equations (2.1). It automatically holds when $Y$ is a scalar and the structural equations have one solution for some values of exogenous variables, while have three solutions for other values of exogenous variables (like the example in Figure 2.3). Such examples can be found in Dagsvik and Jovanovic (1994) and Echenique and Komunjer (2007). Assumption D9(ii) is a stronger version of Assumption D8(i), as the former requires Assumption D8(i) to hold for all $m$ such that $r(A_m) \cap S \neq \varnothing$, and for all $x$. The following Corollary 3 states another version of Proposition 2 when Assumption D6 is replaced by Assumption D7 and D9.

**Corollary 3.** *Suppose the model generates multiple equilibria, and Assumptions D1-D5, D7 and D9 hold. Then there is a set $\mathcal{X}_{D_2} \subset \mathcal{X}$ with* $\Pr(X \in \mathcal{X}_{D_2}) > 0$, *such that for all $x \in \mathcal{X}_{D_2}$ the conclusions of Proposition 2 hold.*

The example in Figure 2.3 satisfies Assumption D9(i). Due to the simplicity of the structural equation, Assumption D9(ii) is sufficient but not necessary for that example to produce a jump in the density when there are multiple equilibria.

Up to now, Proposition 1 and 2 (and Corollaries 2 and 3) all focus on the conditional density $f_{Y|X}(y|x)$. However, a jump in the unconditional density $f_Y(y)$ is usually easier to detect. The next proposition shows that with the additional Assumption D10 below, the relation from uniqueness/multiplicity of equilibria to continuity/discontinuity of the conditional density $f_{Y|X}(y|x)$ can be transferred to the unconditional density $f_Y(y)$.

**Assumption D10.** If the model generates a unique equilibrium, there exists a function $\psi : \cup_{m=1}^{M} B_m \to [0,1]^M$ such that the equilibrium selection rule satisfies $\lambda(x,u) = \psi(g(x)+u)$ for all $(x,u) \in \mathcal{W}$.

When a model is in Category (C1), Assumption D10 holds trivially. Assumption D10 becomes a restriction for a model in (C2) by requiring that the equilibrium selection rule only depends on $g(x)+u$. Using the structural equations (2.1), we can re-write the restriction in Assumption D10 as $\lambda(x,u) = \psi(r(y))$. Assumption D10 is useful in its own right. Recall the example of nonparametric BLP model in Section 2.2, the unique equilibrium assumption imposed by Berry and Haile (2014) corresponded to models with a unique equilibrium satisfying Assumption D10.

Proposition 3 below establishes the relations from the equilibrium behaviour of the model to continuity or discontinuity of the unconditional density of dependent variables.

**Proposition 3**. *Suppose Assumptions D1-D4 hold.*

*(i) Under the additional Assumption D10, if the model generates a unique equilibrium, the unconditional density $f_Y(y)$ is continuous over $y \in int(\mathcal{Y})$.*

*(ii) Under additional Assumptions D5, D7 and D8, (D8 can be replaced by D9), if the model generates multiple equilibria, the unconditional density $f_Y(y)$ has a jump at some $y \in int(\mathcal{Y})$. That is, there exist disjoint open subsets $C_1, C_2 \subset int(\mathcal{Y})$* with *$\bar{C}_1 \cap \bar{C}_2 \neq \varnothing$, and $y_d = \bar{C}_1 \cap \bar{C}_2$ satisfying*

$$\left| \lim_{y \in C_1, y \to y_d} f_Y(y) - \lim_{y \in C_2, y \to y_d} f_Y(y) \right| \quad = \quad \delta > 0.$$

**Remark 3.** (i) Assumption D10 is required for part (i) of Proposition 3 but not for part (ii).

(ii) An intuition for part (ii) of Proposition 3 is that under Assumption D7(i), a jump occurs at the boundary of two subsets of $\bar{\mathcal{Y}}$ such that one of them is associated with a singleton equilibria indices set, while the other is associated with an equilibria indices set containing more than one element. As a result, the signs of the jumps are the same over all $x$'s at which the conditional density $f_{Y|X}(y|x)$ has a jump in $y$. Hence the jump remains in the unconditional density after integrating out $x$ (weighted with its probability).

(iii) Without Assumption D10 (but with other assumptions of Proposition 3 kept), if $f_Y(y)$ is continuous, the underlying model must have a unique equilibrium. On the other hand, if $f_Y(y)$ has a jump, we can further check if $f_{Y|X}(y|x)$ has a jump in $y$ for some $x$ to determine whether the model is in (C2) or (C3).

(iv) Proposition 3 greatly facilitates the implementation of our test when the density jump is estimated nonparametrically. Dimension is reduced substantially by excluding observable characteristics $X$.

In conclusion, we have established a testing criterion that translates testing for multiple equilibria to testing for the presence of a jump in the (conditional) density of the dependent variables.

### 2.3.3 An Extension: Jumps in the Latent Distribution

In this subsection, we consider an extension for our discontinuity criterion discussed so far. Recall that Assumption D4(ii) requires that the the density function of unobservable $U$ is twice continuously differentiable. As a result, if the conditional density $f_{Y|X}(y|x)$ exhibits a jump in $y$, it is due to a jump in the equilibrium selection rule $\lambda(x, u)$, which in turn indicates multiple equilibria. Here we relax Assumption D4(ii) to allow for jumps in the density function $f_U(u)$.

**Assumption D4'.** (i) The unobservable $U$ is independent of $X$. (ii) The distribution of $U$ is absolute continuous with respect to the Lebesgue measure and its density $f_U(u) = dF_U(u)/du$ is bounded and twice continuously differentiable for *almost* all $u \in \mathcal{U}$.

Assumption D4' allows for jump(s) in the density $f_U(u)$, such that the collection of jump locations has a zero measure in $\mathcal{U}$.

Although a jump in $f_U(u)$ will produce a jump in $f_{Y|X}(y|x)$, the following proposition ensures that the jump in $f_{Y|X}(y|x)$ due to jump(s) in $f_U(u)$ will vanish after one integrates out $x$ with its weighted probability, if the distribution of $g(X)$ does not have a mass point.

**Assumption D11.** $\Pr(g(X) \in F_0) = 0$, for any $F_0 \subset \mathbf{R}^J$ with $\mu(F_0) = 0$ (where $\mu$ denotes the Lebesgue measure on $\mathbf{R}^J$).

**Proposition 4.** *Suppose Assumptions D1-D3, D4', D10 and D11 hold. If the model has a unique equilibrium, the unconditional density $f_Y(y)$ is continuous on $y \in int(\mathcal{Y})$.*

The intuition of Proposition 4 as follows. For an arbitrary $y_0$, the jumps in the latent density $f_U(u)$ can produce a jump in the conditional density $f_{Y|X}(y|x)$ at $y_0$, only for $x \in \mathcal{X}_{JL} = \{x : f_U(u) \text{ has a jump at } u = r(y_0) - g(x)\}$. Since $f_U$ has jump(s) at a zero measure set and Assumption D11 holds, we have $\Pr(X \in \mathcal{X}_{JL}) = 0$. Therefore, the jump of conditional density at $y_0$ will disappear in the unconditional density of $Y$, after we integrate out $x$ with its weighted probability. Proposition 3 and 4 together imply that the presence of a jump in the unconditional density $f_Y(y)$ is able to distinguish a jump due to multiple equilibria from that due to jump(s) in the latent density $f_U(u)$.

## 2.4 A Test Statistic

This section proposes a nonparametric test for multiple equilibria via testing for the presence of a jump in the density function of dependent variables. To focus on the main problem, we assume that the assumptions for Proposition 3 hold so that it suffices to consider the unconditional density function $f_Y(y)$. Thus in this section, we can suppress the subscript $Y$ in $f_Y(y)$ without causing confusion. Our testing procedure consists of three steps. Firstly, at a fixed $y$, we compute the local jump of the density. Secondly, we maximize those local jumps over an appropriate set of $y$. This gives our test statistic. Lastly, we compute the critical value by a Gaussian multiplier boostrap. The way to compute the local jump at a fixed $y$ resembles that in Qiu (2002), which focused on estimating the locations of discontinuities in a bivariate regression. The method to compute the critical value is adapted from Chernozhukov, Chetverikov and Kato (2013). We establish asymptotic properties of our test in Proposition 5 and 6. The technical conditions SL1-SL5 for Proposition 5 and 6 are stated in Appendix A.2. The proofs for Proposition 5 and 6 are collected in Appendix A.3.

### 2.4.1 Construction of the Test

**Assumption S1.** $\{Y_i : i = 1, 2, ..., n\}$ is an i.i.d. sample generated by the structural equations (2.1) and an equilibrium selection rule.

Let $\mathcal{J} = \{1, 2, ..., J\}$, by Proposition 3, our testing problem is

$$\begin{aligned} H_0 &: f(y) \text{ is continuous over } y \in \text{int}(\mathcal{Y}). \\ H_1 &: f(y) \text{ has at least a jump within } y \in \text{int}(\mathcal{Y}). \end{aligned}$$

Recall Remark 1, continuity at $y$ means that the limits from all directions coincide. By Assumptions D2, D4 and Lemma 2, the density $f(y)$ is twice continuously differentiable for $y \in \text{int}(\mathcal{Y})$ except the jump(s), if they exist. Let $y_j$ denote the $j$th component of $y$ and $y_{-j}$ denote the components of $y$ other than the $j$th, $f(y_j^+, y_{-j}) = \lim_{\varepsilon \downarrow 0}(y_j + \varepsilon, y_{-j})$ and $f(y_j^-, y_{-j}) = \lim_{\varepsilon \downarrow 0}(y_j - \varepsilon, y_{-j})$. Using this notation, at a fixed $y$, the local density jump along the direction of the $j$th coordinate is

$$\Delta^{(j)}(y) = f(y_j^+, y_{-j}) - f(y_j^-, y_{-j}).$$

By Proposition 3, if the model has a unique equilibrium, $\Delta^{(j)}(y) = 0$ for all $j \in \mathcal{J}$ and $y \in \text{int}(\mathcal{Y})$. On the other hand, if the model has multiple equilibria, $\sup_{j\in\mathcal{J},y\in\mathrm{int}(\mathcal{Y})}\Delta^{(j)}(y)=\delta>0$. We construct a kernel estimator [13] of $\Delta^{(j)}(y)$,

$$\Delta_n^{(j)}(y)=\hat{f}(y_j^+,y_{-j})-\hat{f}(y_j^-,y_{-j}),$$

where $\hat{f}(y_j^+,y_{-j})$ and $\hat{f}(y_j^-,y_{-j})$ are constructed using one-sided kernels defined as follows.

For $j=1,...,J$, let $K_j^+(v_j,v_{-j})$ and $K_j^-(v_j,v_{-j})$ be two non-negative functions which satisfy the following conditions:

(i) The support of $K_j^+(v_j,v_{-j})$ is $[0,1]\times[-1/2,1/2]^{J-1}$ and the support of $K_j^-(v_j,v_{-j})$ is $[-1,0]\times[-1/2,1/2]^{J-1}$.

(ii) $\int_{-1}^1\int_{-1}^1...\int_{-1}^1 K_j^+(v_j,v_{-j})dv=1$ and $\int_{-1}^1\int_{-1}^1...\int_{-1}^1 K_j^-(v_j,v_j)dv=1$.

A typical example is a product kernel. That is, for $j=1,...,J$,

$$\begin{aligned}K_j^+(v_j,v_{-j})&=K^+(v_j)\times\prod_{k\neq j}K(v_k),\\ K_j^-(v_j,v_{-j})&=K^-(v_j)\times\prod_{k\neq j}K(v_k),\end{aligned}\tag{2.13}$$

where $K^+:[0,1]\to\mathbf{R}^+\cup\{0\}$ and $K^-:[-1,0]\to\mathbf{R}^+\cup\{0\}$ are one-sided kernels functions while $K:[-1/2,1/2]\to\mathbf{R}^+\cup\{0\}$ is a usual two-sided kernel function.

Using one-sided kernel functions, we construct estimators of the limits of the density at $y$ and along the $j$th coordinate as

$$\begin{aligned}\hat{f}(y_j^+,y_{-j})&=\frac{1}{np_nh_n^{J-1}}\sum_{i=1}^n K_j^+\left(\frac{Y_{i,j}-y_j}{p_n},\frac{Y_{i,-j}-y_{-j}}{h_n}\right),\\ \hat{f}(y_j^-,y_{-j})&=\frac{1}{np_nh_n^{J-1}}\sum_{i=1}^n K_j^-\left(\frac{Y_{i,j}-y_j}{p_n},\frac{Y_{i,-j}-y_{-j}}{h_n}\right).\end{aligned}\tag{2.14}$$

where $Y_{i,j}$ is the $j$th component of $Y_i$ and $Y_{i,-j}$ is the components of $Y_i$ other than the $j$th. $p_n$ and $h_n$ are bandwidths of the kernel estimator. In

[13] Under $H_0$, $f(y)$ is twice continuously differentiable, hence the kernel estimator of $\Delta^{(j)}(y)$ is consistent for any $y$ and $j$, under standard conditions. Under $H_1$, $f(y)$ is not continuous at some $y$. However, for those $y$, as long as $f(y)$ is twice continuously differentiable over a neighbourhood of $(y_j+\varepsilon,y_{-j})$ and $(y_j-\varepsilon,y_{-j})$ for sufficiently small $\varepsilon$, the one-sided kernel estimator of $\Delta^{(j)}(y)$ remains consistent. Otherwise, the estimator may not be consistent. (for example, if a bivariate density function $f(y_1,y_2)$ has a jump at $y_1=c$, $f(c,y_2)$ is not continuous for all $y_2$). However, this does not affect the level of our test. Moreover, the power of our test will not be reduced if $\sup_{j\in\mathcal{J},y\in\mathrm{int}(\mathcal{Y})}\Delta^{(j)}(y)$ is not underestimated.

addition, define a function $\varphi_n^{(j)} : \mathcal{Y} \times \mathcal{Y} \to \mathbf{R}$ as follows

$$\varphi_n^{(j)}(t,y) = \frac{1}{p_n h_n^{J-1}} \left[ \ K_j^+ \left( \tfrac{t_j - y_j}{p_n}, \tfrac{t_{-j} - y_{-j}}{h_n} \right) - K_j^- \left( \tfrac{t_j - y_j}{p_n}, \tfrac{t_{-j} - y_{-j}}{h_n} \right) \ \right], \tag{2.15}$$

where $t_j$ denotes the $j$th component of $t$ and $t_{-j}$ denotes the components of $t$ other than the $j$th. The same subscript is used for $y$. Therefore, we have

$$\Delta_n^{(j)}(y) = \frac{1}{n} \sum_{i=1}^{n} \varphi_n^{(j)}(Y_i, y).$$

The variance of $\sqrt{n p_n h_n^{J-1}} \Delta_n^{(j)}(y)$ is

$$\sigma_{j,n}^2(y) = p_n h_n^{J-1} \mathbf{E} \left[ \varphi_n^{(j)}(Y_i, y) - \mathbf{E} \left[ \varphi_n^{(j)}(Y_i, y) \right] \right]^2. \tag{2.16}$$

and an estimator of the variance above is

$$\hat{\sigma}_{j,n}^2(y) = \frac{p_n h_n^{J-1}}{n} \sum_{i=1}^{n} \left( \varphi_n^{(j)}(Y_i, y) - \Delta_n^{(j)}(y) \right)^2. \tag{2.17}$$

Therefore, the normalized local jump at the fixed point $y$ and along the fixed direction parallel to the $j$th coordinate can be computed by

$$\left| \sqrt{n p_n h_n^{J-1}} \Delta_n^{(j)}(y) / \hat{\sigma}_{j,n}(y) \right|.$$

Then we consider maximizing local jumps across $J$ coordinates and over an appropriate subset of the support of $Y$. We have to exclude points near the boundary of the support $\mathcal{Y}$ because they tend to distort our test. The local jump may appear large at a point near the boundary of support $\mathcal{Y}$ simply because few sample points are available at one side. In the following, we construct two maximizing sets $\hat{\mathcal{Y}}_{n,\kappa_1}$ and $\hat{\mathcal{Y}}_{n,\kappa_2}$. Firstly for each $y$, we fix a coordinate $j$, and compute the proportion of data points with its $j$th component lying between $y_j - \hat{w}_j$ and $y_j$, conditional on the value of the components other than the $j$th. We denote this proportion as $\hat{P}_j^-(y)$. Here the distance $\hat{w}_j$ is given as $1/H$ of the distance between the largest and the smallest $j$th component of $Y_i$. Similarly, we compute $\hat{P}_j^+(y)$ as the proportion of data points which have the $j$th component lying between $y_j$ and $y_j + \hat{w}_j$, conditional on the value of components other than the $j$th. Then we construct $\hat{\mathcal{Y}}_{n,\kappa_1}$ as the collection of such $y$'s that the ratio of $\hat{P}_j^-(y)$ and $\hat{P}_j^+(y)$ is neither very large nor very small for all coordinates. The cutoff is specified by a constant $\kappa_1$. Analogously, we construct $\hat{\mathcal{Y}}_{n,\kappa_2}$ using the cutoff $\kappa_2$. The data dependent set $\hat{\mathcal{Y}}_{n,\kappa_1}$ serves as a maximizing set for the test statistic while the set $\hat{\mathcal{Y}}_{n,\kappa_2}$ is for the critical value. We set $\kappa_2$ slightly smaller than $\kappa_1$ to facilitate the investigation of asymptotic properties. If the difference between $\kappa_1$ and $\kappa_2$ is small, the power loss due to the discrepancy between $\kappa_1$ and $\kappa_2$ will also be small. In the simulation studies below, we try $(\kappa_1, \kappa_2) = (0.20, 0.19)$ and $(0.25, 0.24)$. Note that our construction of $\hat{P}_j^-(y)$ and $\hat{P}_j^+(y)$ is conditional on the value of components other than the $j$th. This conditioning is necessary when the support of $Y$ is not a rectangle, which is common in the models we consider. Ideally, the choice of $H$ should depend on the shape of the support of $Y$. For example, when the support has a hole (zero density) in the middle, it is better to choose bigger $H$ so that the data points near the boundary of such a hole will be excluded. In our simulation studies below, we set $H = 3$. As described so far, the maximizing sets are constructed as follows.

$$
\begin{aligned}
\hat{\mathcal{Y}}_{n,\kappa_1} &= \{y \in \mathcal{Y} : \hat{P}_j^-(y)/\hat{P}_j(y) > \kappa_1, \hat{P}_j^+(y)/\hat{P}_j(y) > \kappa_1 \text{ all } j\}, \\
\hat{\mathcal{Y}}_{n,\kappa_2} &= \{y \in \mathcal{Y} : \hat{P}_j^-(y)/\hat{P}_j(y) > \kappa_2, \hat{P}_j^+(y)/\hat{P}_j(y) > \kappa_2 \text{ all } j\},
\end{aligned}
$$

where $\kappa_2 < \kappa_1$ are two constants, and

$$
\begin{aligned}
\hat{P}_j^-(y) &= \frac{1}{n}\sum_{i=1}^{n} 1\{y_j - \hat{w}_j \le Y_{i,j} < y_j, |Y_{i,k} - y_k| \le h_n/2, \text{all } k \ne j\}, \\
\hat{P}_j^+(y) &= \frac{1}{n}\sum_{i=1}^{n} 1\{y_j \le Y_{i,j} \le y_j + \hat{w}_j, |Y_{i,k} - y_k| \le h_n/2, \text{all } k \ne j\}, \\
\hat{P}_j(y) &= \hat{P}_j^-(y) + \hat{P}_j^+(y), \ \hat{w}_j = (\max_i Y_{i,j} - \min_i Y_{i,j})/H, \qquad (2.18)
\end{aligned}
$$

for all $j \in \mathcal{J}$.

The next step is to take the supremum of $\left|\sqrt{np_n h_n^{J-1}}\Delta_n^{(j)}(y)/\hat{\sigma}_{j,n}(y)\right|$ across coordinates and over $y \in \hat{\mathcal{Y}}_{j,\kappa_1}$. This gives our test statistic test statistic $\Delta_n$.

$$
\Delta_n = \sup_{j\in\mathcal{J}, y\in\hat{\mathcal{Y}}_{n,\kappa_1}} \left| \frac{\sqrt{np_n h_n^{J-1}}\Delta_n^{(j)}(y)}{\hat{\sigma}_{j,n}(y)} \right|. \qquad (2.19)
$$

The rejection rule is

$$
\text{Reject } H_0 \text{ when } \Delta_n \ge \text{ critical value } \tilde{c}_{1,n}^S.
$$

The process $\{\frac{\sqrt{np_n h_n^{J-1}}(\Delta_n^{(j)}(y)-\mathbf{E}(\Delta_n^{(j)}(y))}{\hat{\sigma}_{j,n}(y)}, y \in \mathcal{Y}\}$ does not converge weakly to a Gaussian process, therefore the calculation of critical value is not standard.[14] Here we use the supremum of the Gaussian multiplier bootstrap proposed by Chernozhukov, Chetverikov and Kato (2013) to simulate critical values. The method consists of three steps:

First draw a sequence of standard normal random variables $\{\xi_1, \xi_2, ..., \xi_n\}$ independent of $\{Y_i\}_{i=1}^n$ and calculate

$$\hat{\mathbb{G}}_{n,j}(y) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \frac{\sqrt{p_n h_n^{J-1}} \left[\varphi_n^{(j)}(Y_i, y) - \Delta_n^{(j)}(y)\right]}{\hat{\sigma}_{j,n}(y)}. \tag{2.20}$$

Then, repeat the first step $R$ times and compute $\hat{c}_{1,n}^S(\alpha)$ as the conditional $(1-\alpha)$th empirical quantile of $\sup_{j\in\mathcal{J}, y\in\hat{\mathcal{Y}}_{n,\kappa_2}} \left|\hat{\mathbb{G}}_{n,j}(y)\right|$.

Lastly, set the critical value as

$$\tilde{c}_{1,n}^S = \hat{c}_{1,n}^S(\alpha) + C\hat{c}_{1,n}(\gamma_n), \tag{2.21}$$

where $1 \geq \gamma_n \to 0$ and $\hat{c}_{1,n}(\gamma_n)$ is the $(1-\gamma_n)$th empirical quantile of

$$\sup_{j\in\mathcal{J}, y\in\hat{\mathcal{Y}}_{n,\kappa_2}} \left|\hat{\mathbb{G}}_{n,j}(y)\right|.$$

The constant $C$ satisfies that $C \geq \bar{M}C_4/((1-\epsilon_{3n})c_2)$. Details of the parameter $(\gamma_n, \bar{M}, C_4\ \epsilon_{3n}, c_2)$ will be specified in Conditions SL2-SL4 (in Appendix A.2), Proposition 5 and Condition SH3 (in Appendix A.3). However, under Conditions SL2, SL4, and SL5 (in Appendix A.2), any positive value $C$ (however small) will give a correct level for our test under the null hypothesis of continuity. Therefore one can set $C = 0$ in practice. The validity of the Gaussian multiplier bootstrap does not depend on the weak convergence of the underlying empirical process, instead it comes from two coupling inequalities (see Theorem 3.1 and 3.2 of Chernozhukov, Chetverikov and Kato (2013)).

[14] In this occasion, the problem of computing the critical value is related to constructing confidence bands of kernel density estimators, which has been studied by Bickel and Rosenblatt (1973), Giné and Nickl (2010), Chernozhukov, Chetverikov and Kato (2013), among others. When $J = 1$, Chu and Cheng (1996) applied the result of Bickel and Rosenblatt (1973) to derive the limit of $\Pr(\sup|\Delta_n(y)| < a_n + b_n z)$ as $\exp(-2\exp(-z))$ for carefully chosen $a_n$ (diverging) and $b_n$ (converging to 0).

### 2.4.2 Asymptotic Properties of the Test

In the section, we describe the asymptotic performance of our test where the statistic is constructed as (2.19) and critical value computed by (2.21). The main results are stated in Proposition 5 and Proposition 6. To obtain those results, we will impose five technical conditions, namely Conditions SL1 to SL5, among which Conditions SL1 to SL4 are adapted from Chernozhukov, Chetverikov and Kato (2013), and Condition SL5 is an under-smoothing condition imposed on the bandwidths $p_n$ and $h_n$ in (2.14). All those technical conditions are stated in Appendix A.2. Here we briefly describe them. Recall the function $\varphi_n^{(j)}$ defined in (2.15), Condition SL1 requires that the following class of functions

$$\Delta\mathcal{K}_{n,j} = \left\{ \frac{\sqrt{p_n h_n^{J-1}}\varphi_n^{(j)}(\cdot, y)}{\sigma_{n,j}(y)} : y \in \mathcal{Y}_{\kappa_2} \right\}$$

is bounded, and its $L_2$ covering number has a specified bound (i.e., is a VC (Vapnik-Cervonenkis) class in the sense of Chernozhukov, Chetverikov and Kato (2013), or is an Euclidean class in the sense of Pakes and Pollard (1989)). Note that here the set $\mathcal{Y}_{\kappa_2}$ is the population version of the maximizing set $\hat{\mathcal{Y}}_{n,\kappa_2}$. The set $\mathcal{Y}_{\kappa_2}$ is specified at the beginning of Appendix A.2. Condition SL2 imposes bounds for the variance given by (2.16). If one uses the product kernel in (2.13), Conditions SL1 and SL2 are satisfied if the kernel functions $K(\cdot)$, $K^{+}(\cdot)$ and $K^{-}(\cdot)$ have a compact support and a bounded variation. Condition SL3 bounds the bias of $\Delta_n^j(y)$. Conditions SL4 and SL5 impose conditions on bandwidths $p_n, h_n$. In this chapter, we let bandwidth $p_n = p_0 \times n^{-\gamma_p}$ and $h_n = h_0 \times n^{-\gamma_h}$. Condition SL5 requires the rate $\gamma_p$ and $\gamma_h$ to satisfy $\gamma_p + (J-1)\gamma_h < 1$, $\gamma_p + (J+1)\gamma_h > 1$, $3\gamma_p + (J-1)\gamma_h > 1$, which is an under-smoothing condition.

The validity of the critical value can be described heuristically as follows. After controlling the difference between $\sigma_{j,n}(y)$ and $\hat{\sigma}_{j,n}(y)$, the desired critical value is approximately the $(1-\alpha)$th quantile of the distribution of $\sup_{j,y\in\hat{\mathcal{Y}}_{n,\kappa_1}} |Z_{j,n}(y)|$ where

$$Z_{j,n}(y) = \frac{\sqrt{np_n h_n^{J-1}}\left(\Delta_n^{(j)} - \mathbf{E}\left[\Delta_n^{(j)}(y)\right]\right)}{\sigma_{j,n}(y)}.$$

Because $\kappa_2 < \kappa_1$, the $(1-\alpha)$th quantile of the distribution of $\sup_{j\in\mathcal{J},y\in\hat{\mathcal{Y}}_{n,\kappa_1}} |Z_{j,n}(y)|$ will be asymptotically bounded by the $(1-\alpha)$th quantile of the distribution of $\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{\kappa_2}}|Z_{j,n}(y)|$. The random variable $\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{\kappa_2}}|Z_{j,n}(y)|$, which is the supremum of an empirical process indexed by a VC class, can be approximated by a random variable with the same distribution as $\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{\kappa_2}}|G_{j,n}(y)|$, where $\{G_{j,n}(y) : j \in \mathcal{J}, y \in \mathcal{Y}_{\kappa_2}\}$ is a Gaussian process with mean zero and variance one (Theorem 3.1 of Chernozhukov et al. (2013)). On the other hand, the supremum of the Gaussian multiplier process in (2.20) can be approximated by a random variable with the same distribution as the random variable $\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{\kappa_2}}|G_{j,n}(y)|$ (Theorem 3.2 of Chernozhukov et al. (2013)). Combining both approximations, the $(1-\alpha)$th quantile of the distribution of $\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{\kappa_2}}|Z_{j,n}(y)|$ can be approximated by the $(1-\alpha)$th quantile of the supremum of the Gaussian multiplier process.

Proposition 5 and 6 below establish the asymptotic behaviour of our test.

**Proposition 5**. *Let the test statistic $\Delta_n$ be constructed as in (2.19) and the critical value $\tilde{c}^S_{1,n}$ be computed as in (2.21). Suppose Assumption S1 and Conditions SL1-SL5 hold. Then, under $H_0$,*

$$\liminf_{n\to\infty} \Pr\left(\Delta_n \le \tilde{c}^S_{1,n}\right) \ge 1-\alpha.$$

The next proposition shows that our test has a power converging to one for any fixed alternative in which a jump occurs not too close to the boundary of the support. Formally, we assume that a jump occurs within the set $\mathcal{Y}_{n,L,c}$ defined as follows.

$$\mathcal{Y}_{n,L,c} = \left\{ \begin{array}{c} y \in \mathcal{Y} : P^-_j(y)/P_j(y) > \kappa_1 + cn^{-d}, \\ P^+_j(y)/P_j(y) > \kappa_1 + cn^{-d},\ \text{all } j \in \mathcal{J} \end{array} \right\},$$

where $P^+_j(y)$, $P^-_j(y)$ and $P_j(y)$ are population versions of $\hat{P}^+_j(y)$, $\hat{P}^-_j(y)$ and $\hat{P}_j(y)$ in (2.18), and are specified at the beginning of Appendix A.2, $c$ is some positive constant and the restriction on the rate $d$ will be specified in Proposition 6.

**Proposition 6**. *Let the test statistic $\Delta_n$ be constructed as in (2.19) and the critical value $\tilde{c}^S_{1,n}$ be computed as in (2.21). Suppose Assumption S1 and Conditions SL1-SL5 hold, and the following conditions are satisfied.*

*(i) The maximal jump in $\mathcal{Y}_{n,L,c}$ is positive, that is,*

$$\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{n,L,c}} \left|\Delta^{(j)}(y)\right| = \delta > 0.$$

*(ii)* $np_n h_n^{J-1}\gamma_n/\log n \to \infty$ *or* $C = 0$, $np_n h_n^{J-1}/\log n \to \infty$. *(*$\gamma_n$ *is in (2.21).)*

*(iii) The rate* $d$ *in* $\mathcal{Y}_{n,L,c}$ *satisfies*

$$0 < d < \min\left(\gamma_h, \gamma_p, (1-\gamma_p-(J-1)\gamma_h)/2\right).$$

*Then*

$$\liminf_{n\to\infty} \Pr\left(\Delta_n > \tilde{c}^S_{1,n}\right) = 1.$$

In this section, we compute the local jump of the density function by checking the jump along each coordinate. A generalization will be to compute the jump along all directions. This can be achieved by using the rotational kernel functions introduced by Qiu (1997). The rotational kernel functions for multi-dimension can be constructed from the spherical coordinate systems[15]. A test based on local jumps that take into account all directions tends to have greater power, however, it is computationally more intensive. We do not go into those details in this chapter.

## 2.5 Monte Carlo Simulations

### 2.5.1 Data Generating Processes (DGP)

The DGPs used in the Monte Carlo simulation are based on Echenique and Komunjer (2007). It is a simplified BLP model but the consumer's preference has an heteroskedastic idiosyncratic term depending on prices. Assume there are two products $j = 1, 2$ produced by two firms. The utility of a consumer is

$$v_{ijt} = -\alpha p_{jt} + g(p_{jt}p_{-jt})\epsilon_{jit}. \tag{2.22}$$

Note that there is no product-specific unobservable. The component $\epsilon_{jit}$ in the idiosyncratic term is i.i.d. across consumers and markets, and has a Gumbel distribution with density $f(t) = \exp(-t + \exp(-t))$. The variance of the idiosyncratic term, however, depends on $p$ through $g(p_j p_{-j})$. Let $g(x) = \rho + \exp(-\tau/x)$. The marginal cost is assumed to be

$$c_{jt} = e^{\gamma}\omega_{jt}, \tag{2.23}$$

[15]The spherical coordinate system for dimensions larger than two can be found in Blumenson (1960).

where $\omega_j$ is product-level unobservable for the cost. The first-order conditions yield the following structural equations in the form of (2.1):

$$\begin{aligned} & \ln\left(p_{1t} - \frac{F(\delta_{1t})g(p_{1t}p_{2t})}{f(\delta_{1t})\alpha}\left[1 + \frac{p_{1t}p_{2t}g'(p_{1t}p_{2t})(p_{2t}-p_{1t})}{g(p_{1t}p_{2t})p_{1t}}\right]^{-1}\right) \\ = \; & \gamma + \ln\omega_{1t}, \\ & \ln\left(p_{2t} - \frac{F(\delta_{2t})g(p_{1t}p_{2t})}{f(\delta_{2t})\alpha}\left[1 + \frac{p_{1t}p_{2t}g'(p_{1t}p_{2t})(p_{1t}-p_{2t})}{g(p_{1t}p_{2t})p_{2t}}\right]^{-1}\right) \\ = \; & \gamma + \ln\omega_{2t}, \end{aligned} \tag{2.24}$$

where $\delta_{jt} = \alpha(p_{-jt} - p_{jt})/g(p_{jt}p_{-jt})$ for $j = 1, 2$, and an equilibrium in market $t$ is $(p_{1t}, p_{2t})$. Function $F(t) = 1/(1 + \exp(-t))$ is the CDF of the Logistic distribution. The number of solutions to equations (2.24) depends on values of parameter $(\alpha, \gamma, \rho, \tau)$ and $(\omega_1, \omega_2)$. Table 2.2 and Figure 2.4 show the cases of a unique and three solutions when fixing $(\omega_1, \omega_2)$ at $(1, 1)$. The value of parameters $(\alpha, \gamma, \rho, \tau)$ follows that in Echenique and Komunjer (2007). In Table 2.2 , the left column computes the three equilibrium prices for the structural equations (2.24), with parameters as specified in the table and $(\omega_1, \omega_2) = (1, 1)$. The right column, on the other hand, computes the unique equilibrium price for the the structural equations (2.24), with parameters as specified in the table and $(\omega_1, \omega_2) = (1, 1)$. In Figure 2.4, the upper graph plots the case of multiple solutions while the lower graph plots the case of a unique solution. In each graph, the blue line is the equilibria characterizing function derived as follows: we first solve the best response function for $p_2$ in terms of $p_1$ from the second equation of (2.24), and then plug the best response function into the first equation of (2.24). This way we obtain an equation charactering the equilibrium $p_1$, and then write it as a function equal zero. Such a function is the equilibria characterizing function plotted in Figure 2.4. The red line is the zero horizontal line. The intersection of the red and blue line gives the equilibrium $p_1$. Since equations (2.24) are symmetric when $(\omega_1, \omega_2) = (1, 1)$, we have $p_2 = p_1$ in every equilibrium.

In the Monte Carlo study, we consider the following four DGPs:

$$\begin{aligned} DGP0 \quad &: \quad \alpha = 0.3841, \gamma = 0.0923, \rho = 0.1755, \tau = 11.0009 \\ \omega_1 \quad &\sim \quad U[0.8, 1], \omega_2 \sim U[0.6, 2.5] \\ DGP1,2,3 \quad &: \quad \alpha = 0.2464, \gamma = 0.0776, \;\rho = 0.1074, \tau = 12.9913 \end{aligned}$$

Figure 2.4: Characterization of equilibria

$\alpha = 0.2464, \gamma = 0.0776, \rho = 0.1074, \tau = 12.9913$


$\alpha = 0.3841, \gamma = 0.0923, \rho = 0.1755, \tau = 11.0009$

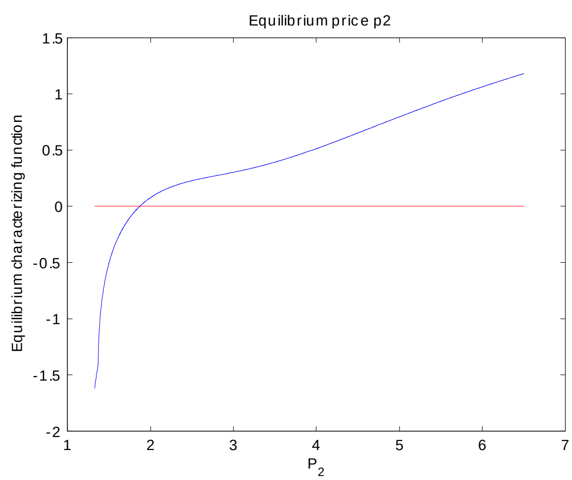


Table 2.2: Equilibrium prices

| $\alpha = 0.2464, \gamma = 0.0776$ <br> $\rho = 0.1074, \tau = 12.9913$ | $\alpha = 0.3841, \gamma = 0.0923$ <br> $\rho = 0.1755, \tau = 11.0009$ |
|---|---|
| $(p_1, p_2)$ | $(p_1, p_2)$ |
| $(1.753, 1.753)$ | $(1.853, 1.853)$ |
| $(3.098, 3.098)$ | This is the unique |
| $(4.843, 4.843)$ | equilibrium |

$$\omega_1 \quad \sim \quad U[0.8, 1], \omega_2 \sim U[0.6, 1.5]$$

The support of $\omega_2$ is larger in DGP0 so that the generated data $(P_1, P_2)$ has a comparable range with the other three. For DGP0, equations (2.24) always admit a unique solution. On the other hand, DGP1, DGP2 and DGP3 generate three solutions with a probability around 0.46 and a unique solution with a probability around 0.54. The difference among DGP1, DGP2 and DGP3 lies in the equilibrium selection rule. We assume that an equilibrium is randomly selected given the equilibria indices set. When the equilibria indices set contains equilibrium $1, 2$ and 3, the equilibrium selection rule of DGP1 always chooses equilibrium 3, that is, DGP1 falls into category (C2) and the model has a unique equilibrium. The equilibrium selection rule of DGP2 chooses equilibrium 1 and 3 with probability 1/4 and 3/4 respectively. The equilibrium selection rule of DGP3 assigns equal probabilities to equilibrium $1, 2$ and 3. As a result, DGP2 and DGP3 fall into category (C3) and produce multiple equilibria. The following table summarizes the equilibrium selection rules.

| DGP | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| # of solutions to (2.24) | 1 | 1 or 3 | 1 or 3 | 1 or 3 |
| Equilibrium selection rule | [1] | $[0, 0, 1]$ | $[1/4, 0, 3/4]$ | $[1/3, 1/3, 1/3]$ |
| # of realized equilibria | 1 | 1 | 1 or 2 | 1 or 3 |

In each time of simulation, the data is an i.i.d. sample $(P_1, P_2)$. Figure 2.5 plots four samples generated by DGP0 to DGP3. Note that the data with a unique equilibrium may have a connected support (like DGP0), however, it may not (like DGP1). Similarly, the data with multiple equilibria may (like DGP3) or may not (like DGP2) have a connected support.

Figure 2.5: Data generated by DGP0,1,2,3

DGP0 (upper left),DGP1 (upper right),
DGP2 (lower left) and DGP3 (lower right)

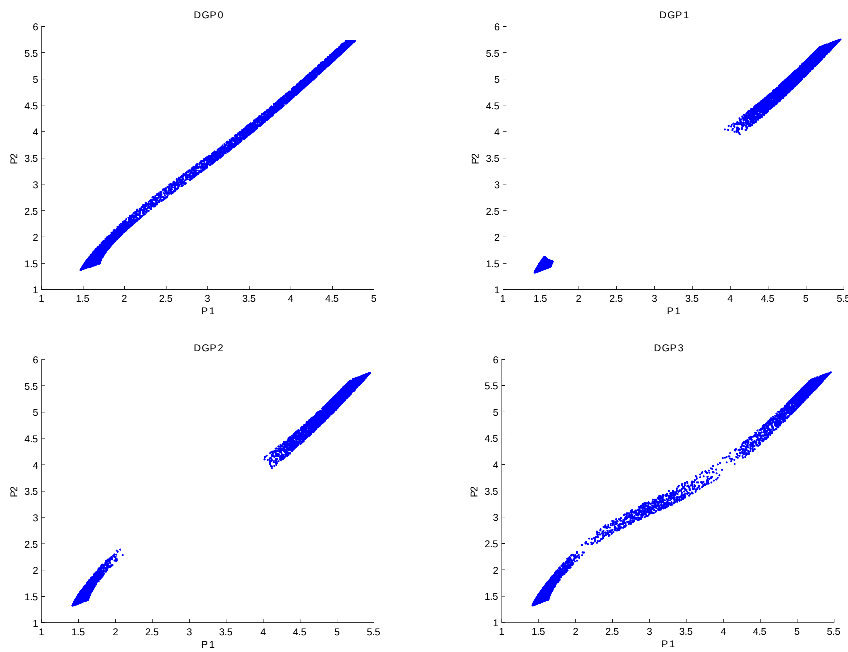


Figure 2.6 magnifies the data generated by DGP2 and DGP3 near the locations of jump. We clearly see that there is a line separating the graph into a dense part and a sparse part. Such a line is the jump location curve, which is the collection of loci where jumps in the density occur.

### 2.5.2 Performance of the Test

The data is the prices $\{P_1, P_2\}$. We compute the test statistic $\Delta_n$ as in (2.19) and the critical value $\tilde{c}^S_{1,n}$ as in (2.21), with $C = 0$. The number of simulation is $R = 500$. The number of simulation used to compute the critical value is $B = 199$. Condition SL5 implies that the bandwidths $p_n = p_0 \times n^{-\gamma_p}$ and $h_n = h_0 \times n^{-\gamma_h}$ must satisfy

$$\gamma_p + \gamma_h < 1, 3\gamma_p + \gamma_h > 1, \gamma_p + 3\gamma_h > 1.$$

Here we use $\gamma_p = \gamma_h = 0.3$. In the construction of $\Delta_n^{(j)}(y)$, $j = 1, 2$, we allow $p_0$ and $h_0$ to be data dependent by letting $p_0 = p_c \times\ \hat{\sigma}_j$ and $h_0 = h_c \times \hat{\sigma}_{-j}$, where $\hat{\sigma}_j$ and $\hat{\sigma}_{-j}$ are the standard deviation of $P_j$ and $P_{-j}$ respectively, $j = 1, 2$. We use the following one-sided Epanechnikov kernels[16],

$$\begin{aligned}
K_1^+(u_1, u_2) &= \frac{12}{11}(1 - u_2^2)1\{|u_2| \le 0.5\} \cdot \frac{12}{11}(1 - (u_1 - 0.5)^2)1\{0 \le u_1 \le 1\}, \\
K_1^-(u_1, u_2) &= \frac{12}{11}(1 - u_2^2)1\{|u_2| \le 0.5\} \cdot \frac{12}{11}(1 - (u_1 + 0.5)^2)1\{-1 \le u_1 \le 0\}, \\
K_2^+(u_1, u_2) &= \frac{12}{11}(1 - u_1^2)1\{|u_1| \le 0.5\} \cdot \frac{12}{11}(1 - (u_2 - 0.5)^2)1\{0 \le u_2 \le 1\}, \\
K_2^-(u_1, u_2) &= \frac{12}{11}(1 - u_1^2)1\{|u_1| \le 0.5\} \cdot \frac{12}{11}(1 - (u_2 + 0.5)^2)1\{-1 \le u_2 \le 0\}.
\end{aligned}$$

Table 2.3 reports the rejection frequencies of our test under different DGPs with parameter $(p_c, h_c) = (0.25, 0.25)$, $(\kappa_1, \kappa_2) = (0.20, 0.19)$ and $H = 3$. The first two rows show the level of our test under DGP0 and DGP1. The last two rows show the power of the test under DGP2 and DGP3. The reported rejection frequencies suggest that the test has a desirable level when there is a unique equilibrium (DGP0 and DGP1) and it also exhibits a reasonable power to detect multiple equilibria in models generated by DGP2 and DGP3. Table 2.4 shows the rejection frequencies for different choices of $(p_c, h_c)$, when $(\kappa_1, \kappa_2) = (0.20, 0.19)$ and sample size $N = 2000$. Table 2.5 uses a different $(\kappa_1, \kappa_2) = (0.25, 0.24)$. Overall, our test performs reasonably well under the specified set of tuning parameters.

[16] Qiu (2002) uses them for a local discontinuity test.

Figure 2.6: Jump location curves
DGP2 (upper) and DGP3 (down)

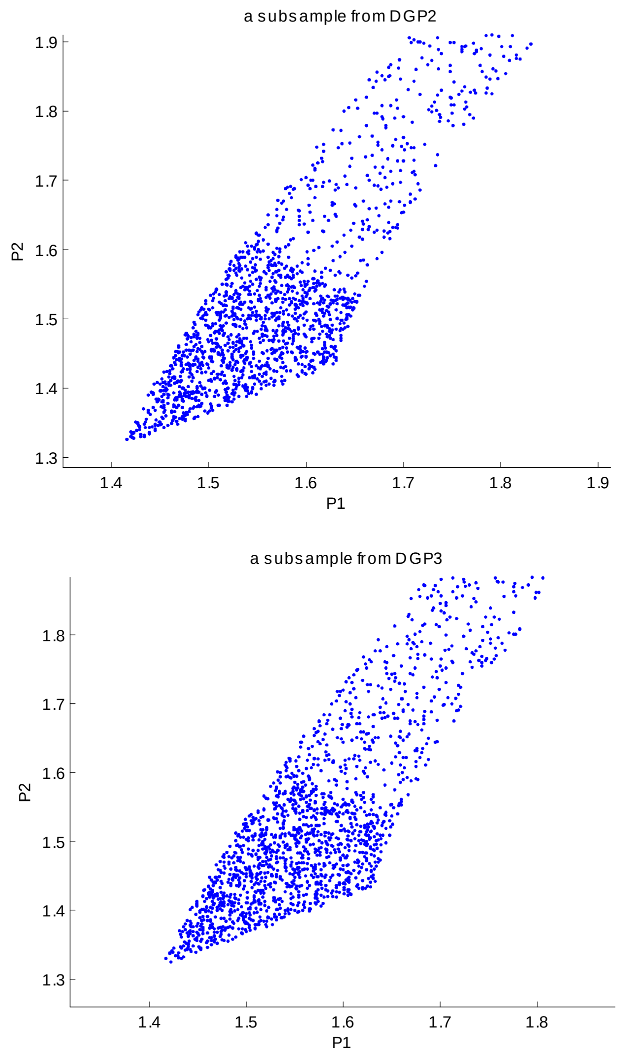


Table 2.3: Rejection frequencies of the test

| sample size | 1500 | | 2000 | |
|---|---|---|---|---|
| Unique equilibrium | | | | |
| nominal $\alpha$ | $DGP0$ | $DGP1$ | $DGP0$ | $DGP1$ |
| 0.05 | 0 | 0.016 | 0 | 0.032 |
| 0.10 | 0.002 | 0.084 | 0.008 | 0.072 |
| Multiple equilibria | | | | |
| | $DGP2$ | $DGP3$ | $DGP2$ | $DGP3$ |
| 0.05 | 0.502 | 0.366 | 0.602 | 0.414 |
| 0.10 | 0.632 | 0.492 | 0.732 | 0.552 |

Table 2.4: Rejection frequencies under different bandwidths

| $(p_c, h_c)$ | $0.15, 0.6$ | | $0.2, 0.45$ | | $0.3, 0.15$ | | $0.35, 0.1$ | | $0.5, 0.06$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Unique equilibrium | | | | | | | | | |
| nominal $\alpha$ | $DGP0$ | $DGP1$ | $DGP0$ | $DGP1$ | $DGP0$ | $DGP1$ | $DGP0$ | $DGP1$ | $DGP0$ | $DGP1$ |
| 0.05 | 0 | 0.056 | 0.006 | 0.062 | 0 | 0.068 | 0 | 0.044 | 0 | 0.064 |
| 0.10 | 0.08 | 0.092 | 0.013 | 0.126 | 0 | 0.136 | 0 | 0.102 | 0 | 0.142 |
| | Multiple equilibria | | | | | | | | | |
| | $DGP2$ | $DGP3$ | $DGP2$ | $DGP3$ | $DGP2$ | $DGP3$ | $DGP2$ | $DGP3$ | $DGP2$ | $DGP3$ |
| 0.05 | 0.236 | 0.194 | 0.598 | 0.402 | 0.428 | 0.344 | 0.292 | 0.248 | 0.354 | 0.226 |
| 0.10 | 0.344 | 0.296 | 0.704 | 0.548 | 0.578 | 0.456 | 0.460 | 0.396 | 0.582 | 0.406 |

Table 2.5: Rejection frequencies under different maximizing sets

$\kappa_1 = 0.25, \kappa_2 = 0.24$

| $(p_c, h_c)$ | $(0.25, 0.25)$ | | $(0.2, 0.45)$ | |
|---|---|---|---|---|
| Unique equilibrium | | | | |
| nominal $\alpha$ | $DGP0$ | $DGP1$ | $DGP0$ | $DGP1$ |
| 0.05 | 0.004 | 0.018 | 0.012 | 0.005 |
| 0.10 | 0.004 | 0.046 | 0.012 | 0.044 |
| Multiple equilibria | | | | |
| | $DGP2$ | $DGP3$ | $DGP2$ | $DGP3$ |
| 0.05 | 0.228 | 0.280 | 0.396 | 0.304 |
| 0.10 | 0.304 | 0.364 | 0.492 | 0.442 |

Our test involves calculating the maximal local jump over values of dependent variables and across the coordinates. Computation of the maximizer may take some time when implementing this test. In our simulation study, the average CPU time (per simulation) for results in the first two rows of Table 2.3 (sample size is 1500) is about 10 to 11 seconds. (In particular, DGP0 10.4 seconds, DGP1 10.3 seconds, DGP2 10.5 seconds and DGP3 10.7 seconds). In our simulations, the dependent variables (the prices for products 1 and 2) are of two dimensions. We expect the computation time for implementing our test to increase when the dimension of the dependent variables becomes higher.

## 2.6 Conclusions and Remarks

This chapter makes the first attempt to propose a test for multiple equilibria when the dependent variables are continuous and the structural equations are nonparametric. We show that uniqueness or multiplicity of equilibria produces testable implications on the continuity or discontinuity of the (conditional) density function of the dependent variables. Based on that, we develop a test for multiple equilibria by testing for the presence of jumps in a multivariate density function. The test statistic is constructed as the supremum of the local density jumps, and the critical value is computed via the Gaussian multiplier bootstrap. Monte Carlo studies show that the test performs reasonably well under different DGPs. Focusing on continuous dependent variables, our test complements the recent literature of testing multiple equilibria in discrete games.

We are aware of other potential testing criteria for multiple equilibria in continuous dependent variables. For example, Berry and Haile (2013) showed that the derivative $\nabla r(y)$ in (2.1) is over-identified if the model is in Category (C1). If we are agnostic about the equilibria behaviour, the over-identification restrictions can potentially serve as a testing criterion for $\nabla_x \lambda(x, u) = 0$ for almost all $x$, which is closely related to the equilibrium behaviour of the model (2.1). A detailed analysis of a test using over-identification restrictions is beyond the scope of this chapter. It is worth noticing that a potential test for multiple equilibria based on over-identification restriction is a complement rather than a substitute for the test via jump(s) in the density proposed in this chapter. The main reason is that the over-identification restrictions will be functions of conditional density $f_{Y|X}(y|x)$, whose estimation is much easier if we know the presence and location of jump(s).

# Chapter 3

# Jumps of the Conditional Choice Probabilities in Incomplete Information Games

## 3.1 Introduction

Research on identification and estimation of simultaneous games with incomplete information has been growing in the past two decades. Related literature includes Sweeting (2009), Bajari, Hong, Krainer and Nekipelov (2010), Aradillas-Lopez (2010), Wan and Xu (2014), Aguirregabiria and Mira (2013), among others. In most of the literature, the conditional choice probabilities are the main object the econometrician can observe. [17] Furthermore, these conditional choice probabilities are usually the starting point for parameter identification and estimation (see for example, Bajari, Hong, Krainer and Nekipelov (2010), Paula and Tang (2012), Wan and Xu (2014), Aguirregabiria and Mira (2013), among others). If all covariates are discrete, the conditional choice probabilities can be nonparametrically estimated using sub-sample means. However, this chapter shows that the conditional choice probabilities may have jump(s) with respect to the continuous covariates (if they exist), even when the payoff functions and distribution of latent variables are all smooth. The source of such jumps lies in the equilibrium behaviour of the game. As a result, it may be too strong to impose the standard smoothness condition on the conditional choice probabilities and the standard nonparametric techniques (kernel

[17] To be precise, in this chapter, when the private information can be correlated across players, a conditional choice probability means the probability of a particular player's choice, conditional on the choices of other players as well as the value of covariates. As a special case, when the private information is independent across players, the conditioning set contains only covariates.

type estimators or series estimators) may not apply immediately. If the econometrician knows that there must be at least a jump in the conditional probability, she/he can first estimate the jump location by the method of Müller (1992) and Delgado and Hidalgo (2000). However, for incomplete information games, the econometrician typically knows *neither* the presence *nor* the location of the jump in the conditional choice probabilities. Therefore, a test for the presence of a jump in the conditional choice probabilities would help the econometrician in choosing appropriate estimation approaches.

Apart from the statistical concern due to the possibility of a jump, the presence of a jump in the conditional choice probabilities provides information about the equilibrium behaviour of the game. As this chapter will show, when the equilibrium characterizing equations always admit a unique solution, the conditional choice probabilities will be continuous on the support of continuous covariates. On the other hand, if there are multiple equilibria, or the single equilibrium present in the data varies in type over the support of covariates, the conditional choice probabilities have a jump, under some reasonable conditions. Such a relationship between the equilibrium behaviour and the presence of a jump in the conditional choice probabilities is robust to correlated private information and unobserved heterogeneity independent of covariates.

Uniqueness or multiplicity of equilibria is important in game-theoretic econometric models. For games with incomplete information, the presence of multiple equilibria affects various stages of the empirical research, such as identification, estimation, comparative static and conterfactual experiments. Aradillas-Lopez (2010) and Wan and Xu (2014) maintained the unique equilibrium assumption for their identification and estimation strategies in incomplete information games with correlated private information.

### 3.1.1 Related Literature

Aguirregabiria and Mira (2013) noticed that the conditional choice probability may have a jump with respect to covariates in incomplete information games with independent private information. This chapter elaborates on the theoretical foundation of the jumps in the conditional choice probabilities. Firstly, the game set-up in this chapter allows for the correlated private information among players. Secondly, we explicitly state the conditions under which the conditional choice probabilities have a jump, and we also relate the presence of jump(s) to the equilibrium behaviour of the game. As a result, the jump is more than a complication for estimation of the conditional choice probabilities, it also reveals information about the equilibrium behaviour of the game. Thirdly, we emphasize that the econometrician is typically agnostic about the presence of a jump. Aguirregabiria and Mira (2013) suggested using the jump location estimator developed by Müller (1992) and Delgado and Hidalgo (2000) when there are jumps. However, the validity of that method relies on the existence of a jump. When there is actually no jump, the aforementioned method no longer applies. Hence we recommend a test for the presence of a jump at the first stage.

When the private information is independent among players within the game, a test for multiple equilibria has been proposed by Paula and Tang (2012). They exploited the equivalence between the uniqueness of equilibrium and the conditional independence among players' actions. Unfortunately, such an equivalence breaks down for correlated private information. In this case, even a unique equilibrium can yield correlation among players' actions. Our result, which relates the equilibrium behaviour to the presence of a jump, is robust to correlated private information across players, although it is weaker than that of Paula and Tang (2012) when players' private information is indeed independent.

This chapter is also related to the literature on the identification and estimation of incomplete information games, because the equilibrium behaviour may affect the identification and estimation approaches. When the private information is independent across players, identification and estimation of incomplete information games have been studied by Bajari, Hong, Krainer and Nekipelov (2010), Aguirregabiria and Mira (2013), Xiao (2014) and many others. Not much research has been conducted for games with correlated private information. Aradillas-Lopez (2010), Wan and Xu (2014) are such examples, and both of them assumed the uniqueness of equilibrium.

A test for the presence of a jump in the conditional choice probabilities (or generally, for a regression function) usually involves a supremum of an empirical process which does not weakly converge. This problem, especially for a univariate regression function (or a density function) has been treated by Bickel and Rosenblatt (1973), Johnston (1982), Chu and Wu (1994), Hamrouni (1999), among others. These papers all established a Gumbel type limiting distribution. A recent development on this issue can be found in Chernozhukov, Chetverikov and Kato (2012, 2013) and Chernozhukov, Lee and Rosen (2013). Qiu (1997, 2002) provided practical methods to detect the jump of a bivariate regression function at a fixed point.

### 3.1.2 Organization of Chapter 3

The organization of this chapter is as follows, Section 3.2 sets up the framework of games with incomplete information; Section 3.3 investigates the implications of the equilibrium behaviour of the game for continuity or discontinuity of the conditional choice probabilities. A numerical example illustrates such relations. We also allow for heterogeneity in the payoff functions unobserved by the econometrician (but independent of the covariates). Section 3.4 points out an econometric consequence of not knowing the presence of a jump in the conditional choice probabilities. Section 3.5 briefly discusses possible methods for testing the presence of a jump in the conditional choice probabilities. The last section concludes. All proofs are collected in Appendix B.2.

## 3.2 The Set-up of the Game

We consider a simultaneous game with incomplete information. To illustrate the main idea, let us focus on games with two players. Let $k = 1, 2$ denotes the identity of the players. Each player chooses an action $Y_k \in \{0, 1\}$. For $k = 1, 2$, the payoff function for player $k$ can be written as

$$\pi_k(Y_k, Y_{-k}, X_k) - U_k(Y_k), \tag{3.1}$$

where $Y_{-k}$ represents the choice of the player other than $k$, $X_k$ is a vector of exogenous covariates for player $k$'s payoff function, and $U_k(Y_k)$ is the random shock to player $k$'s payoff when she chooses action $Y_k$. Without loss of generality, we normalize $U_k(0) = 0$ for all $k$ and simplify the notation of $U_k(1)$ to $U_k$. We summarize the payoff of the game in Table 3.1. The information structure is as follows. The shock $U_k$ to player $k$'s payoff is private information only observed by the player $k$. The joint CDF of $(U_1, U_2)$, denoted as $F_{1,2}(u_1, u_2)$, the functional form $\{\pi_1(\cdot), \pi_2(\cdot)\}$, and the exogenous covariates $X = (X_1, X_2)$ are common knowledge for both players. Denote the support of $X$ as $\mathcal{X}$. The game is played $n$ times and the econometrician observes a random sample $\{(Y_{1i}, Y_{2i}, X'_{1i}, X'_{2i})\}_{i=1}^n$.

We impose two assumptions on the payoff functions and the information structure of the game.

**Assumption 1.** (i) The distribution of covariates $X = (X_1, X_2)$ is absolutely continuous with respect to the Lebesgue measure and its density $f_X$ is bounded and twice continuously differentiable.

Table 3.1: Normal-form of the game

| | $y_2 = 0$ | $y_2 = 1$ |
|---|---|---|
| $y_1 = 0$ | $0, 0$ | $0, \pi_2(1, 0, x_2) - u_2$ |
| $y_1 = 1$ | $\pi_1(1, 0, x_1) - u_1, 0$ | $\pi_1(1, 1, x_1) - u_1, \pi_2(1, 1, x_2) - u_2$ |

(ii) The function $\pi_k(y_k, y_{-k}, x_k)$ is twice continuously differentiable with respect to $x_k$, for $k = 1, 2$.

(iii) The covariates $X$ are observed by players and the econometrician.

**Assumption 2.** (i) The random vector $U = (U_1, U_2) \in \mathbf{R}^2$ is continuously distributed with a joint CDF $F_{1,2}(\cdot)$, which is twice continuously differentiable.

(ii) $U$ is independent of $X$.

(iii) The support of $(U_1, U_2)$ satisfies

$$\begin{aligned} \inf \mathcal{U}_1 &< \inf_{x \in \mathcal{X}} \min \left( \pi_1(1, 0, x_1), \pi_1(1, 0, x_1) + \Delta_1(x_1) \right), \\ \inf \mathcal{U}_2 &< \inf_{x \in \mathcal{X}} \min \left( \pi_2(1, 0, x_2), \pi_2(1, 0, x_2) + \Delta_2(x_2) \right). \end{aligned}$$

where $\Delta_k(x_k) = \pi_k(1, 1, x_k) - \pi_k(1, 0, x_k)$ for $k = 1, 2$.

Assumption 2 (iii) is sufficient for the structural functions $\varphi(\sigma, x)$ in equations (3.7) defined below to be twice continuously differentiable.

In the following, we will characterize the equilibrium of the game under Assumptions 1 and 2. Wan and Xu (2014) pointed out that two notions of equilibrium have been utilized in game-theoretic econometric models with correlated private information. In the first notion, an equilibrium is a pair of self-consistent beliefs $(\sigma_1(1|1, x), \sigma_2(1|1, x))'$ satisfying equations (3.2), where $\sigma_k\,(1|1, x)$ is the player $-k$'s belief of player $k$ choosing action 1, conditional on player $-k$ herself choosing action 1 and the covariate value $x$. The players' optimal actions are determined by

$$\begin{aligned} Y_{1i} &= 1 \left\{ v_1(1, X_i) - U_{1i} > 0 \right\}, \\ Y_{2i} &= 1 \left\{ v_2(1, X_i) - U_{2i} > 0 \right\}, \end{aligned} \tag{3.2}$$

where

$$v_1(1, x) = \pi_1(1, 0, x_1) + \sigma_2(1|1, x)\Delta_1(x_1),$$

$$v_2(1,x) \quad = \quad \pi_2(1,0,x_2) + \sigma_1(1|1,x)\Delta_2(x_2), \tag{3.3}$$

are the expected payoffs of player $k$ conditional on herself choosing action 1 and the realization of covariates $X$, for $k = 1,2$. Aradillas-Lopez (2010) used this notion of equilibrium.

In the second notion, an equilibrium is defined as a pair of cut-offs in the value of private information $(u_1^*(x), u_2^*(x))$, such that the following conditions hold:

$$Y_{k,i} = 1\left\{U_{ki} \le u_k^*(X_i)\right\}, \text{ for } k = 1,2,$$

and

$$\begin{aligned} \pi_1(1,0,x_1) + \Delta_1(x_1)\tilde{\sigma}_2(x) &= u_1^*(x), \\ \pi_2(1,0,x_2) + \Delta_2(x_2)\tilde{\sigma}_1(x) &= u_2^*(x), \end{aligned} \tag{3.4}$$

where

$$\tilde{\sigma}_k(x) = \Pr\left(U_{ki} \le u_k^*(x) | X_i = x, U_{-ki} = u_{-k}^*(x)\right).$$

Note that the existence of a monotone pure Nash equilibrium is assumed. The differences between two notions are as follows. (i) In the first notion, the belief of the other player's action is conditional on the player's own action, whereas in the second one, it is conditional on the player's own private information. (ii) In the second notion, there is an additional assumption of the monotonicity of the equilibrium. In the main text of this chapter, we will follow Aradillas-Lopez (2010) and use the first notion of equilibrium. In Appendix B.1, we investigate the outcomes under the second notion of equilibrium. We will see that all the results obtained under the first notion remain valid under the second notion, with a slight modification.

Now we characterize the equilibrium of the first notion, following Aradillas-Lopez (2010). Given a realization of $X$, an equilibrium is a vector

$$(\sigma_1(1|1,x), \sigma_2(1|1,x))'$$

satisfying the following system of equations.

$$\begin{aligned} \sigma_k(1|1,x) &= \frac{\Pr\left(v_1(1,x) > U_{1i}, v_2(1,x) > U_{2i} | X = x\right)}{\Pr\left(v_{-k}(1,x) > U_{-ki} | X = x\right)} \\ &= \frac{F_{1,2}\left(v_1(1,x), v_2(1,x)\right)}{F_{-k}\left(v_{-k}(1,x)\right)}, \end{aligned} \tag{3.5}$$

for $k = 1,2$, where $F_1(\cdot), F_2(\cdot)$ are the marginal distribution functions of $U_1, U_2$. Note that the first equality in (3.5) comes from (3.2) and the second uses Assumption 2 (ii). In other words, the system of equations characterizing equilibrium beliefs $(\sigma_1(1|1,x), \sigma_2(1|1,x))'$ can be written as

$$\begin{bmatrix} \sigma_1(1|1,x) \\ \sigma_2(1|1,x) \end{bmatrix} = \begin{bmatrix} \frac{F_{1,2}(v_1(1,x),v_2(1,x))}{F_2(v_2(1,x))} \\ \frac{F_{1,2}(v_1(1,x),v_2(1,x))}{F_1(v_1(1,x))} \end{bmatrix}. \tag{3.6}$$

To save on notation, define

$$\varphi(\sigma,x) \equiv \begin{bmatrix} \varphi_1(\sigma,x) \\ \varphi_2(\sigma,x) \end{bmatrix} \equiv \begin{bmatrix} \sigma_1 - \frac{F_{1,2}(\pi_1(1,0,x_1)+\sigma_2\Delta_1(x_1),\pi_2(1,0,x_2)+\sigma_1\Delta_2(x_2))}{F_2(\pi_2(1,0,x_2)+\sigma_1\Delta_2(x_2))} \\ \sigma_2 - \frac{F_{1,2}(\pi_1(1,0,x_1)+\sigma_2\Delta_1(x_1),\pi_2(1,0,x_2)+\sigma_1\Delta_2(x_2))}{F_1(\pi_1(1,0,x_1)+\sigma_2\Delta_1(x_1))} \end{bmatrix}.$$

The system of structural equations can be stated as

$$\varphi(\sigma,x) = 0. \tag{3.7}$$

Note that it follows from Assumptions 1 and 2 that the function $\varphi : [0,1]^2 \times \mathcal{X} \to \mathbf{R}^2$ is twice continuously differentiable.

## 3.3 An Implication of Equilibrium Behaviour on Data

The observed data for the econometrician can be summarized by the conditional choice probabilities defined as

$$\begin{aligned} Q_k\left(y_k|y_{-k},x\right) &\equiv \Pr\left(Y_k = y_k|Y_{-k} = y_{-k}, X = x\right), \\ \text{for } k &= 1,2, \text{ and } (y_1,y_2) \in \{0,1\}^2. \end{aligned}$$

Let $Q\left(1|1,x\right) = \left(Q_1\left(1|1,x\right), Q_2\left(1|1,x\right)\right)'$. In this section, we analyse the implications of the equilibrium behaviour of the game for continuity or discontinuity of the conditional choice probabilities $Q\left(1|1,x\right)$. We begin with studying the relation from $x$ to the equilibrium belief $\sigma$. When (3.7) admits multiple solutions, the relation from the $x$ to $\sigma$ can be a correspondence. Each function $q_m$ in the following definition can be viewed as a branch of this correspondence.

**Definition 1.** *Let $M$ be the smallest number such that there exists a family of twice continuously differentiable functions*

$$\left\{q_1(\cdot), q_2(\cdot), ..., q_M(\cdot)\right\},$$

*where $q_m : B_m \to [0,1]^2$, $B_m$ is an open and connected subset of $\mathcal{X}$ ,and the following conditions are satisfied*:

*(i) For any $(\sigma, x) \in [0,1]^2 \times B_m$ satisfying $\varphi(\sigma, x) = 0$, we have $\sigma = q_m(x)$. Furthermore, $\varphi(q_m(x), x) = 0$ for all $x \in B_m$.*

*(ii) $\cup_{m=1}^M \bar{B}_m = \mathcal{X}$. (Here $\bar{B_m}$ denotes the closure of $B_m$.)*

**Remark 1.** Without loss of generality, we can further assume that for any $x \in B_m \cap B_k$ and $m \neq k$, we have $q_m(x) \neq q_k(x)$. Otherwise we can exclude $x$ from either $B_m$ or $B_k$.

For the numerical example in Section 3.3.2, Figure 3.2 plots the family of functions $\{q_1, q_2, q_3\}$.

The following assumption is imposed on the constant $M$ and the structural equations $\varphi(\sigma, x)$.

**Assumption 3.** (i) The system of structural equations (3.7) has at least one solution for any $x \in \mathcal{X}$.

(ii) The constant $M$ in Definition 1 is finite.

(iii) If the system of structural equations (3.7) admits a unique solution in $\sigma$ for all $x \in \mathcal{X}$, the partial derivative $\bigtriangledown_\sigma \varphi(\sigma, x)$ is invertible for all $(\sigma, x) \in (0,1)^2 \times$ int$(\mathcal{X})$.

Assumption 3 (i) and (ii) require that a solution to (3.7) must exist, and the number of solutions is finite. Assumption 3 (iii) is a technical condition that ensures the partial derivative of $\varphi(\sigma, x)$ with respect to $\sigma$ be invertible in the case of unique solution. Definition 1 yields an index system with regard to the "types" of equilibrium. For each $x$, if a solution $\sigma(x)$ to (3.7)belongs to $q_m(B_m)$, we call it type $m$ equilibrium and denote it as $\sigma^{(m)}(1|1, x) = (\sigma_1^{(m)}(1|1, x), \sigma_2^{(m)}(1|1, x))'$. If $\sigma \in q(B_m)$ for multiple $m$'s, then we take the smallest one as its index. Thus, we have $\sigma^{(m)}(1|1, x) = q_m(x)$ for all $x \in \cup_{m=1}^M B_m$. We define $\Upsilon(x)$ as the equilibria indices set for a generic value $x \in \mathcal{X}$,

$$\Upsilon(x) \equiv \{m : \varphi(q_m(x), x) = 0\}. \tag{3.8}$$

By definition, $\sigma^{(m)}(1|1, x)$ exists if and only if $m \in \Upsilon(x)$. Furthermore, we define an equilibrium selection rule $\pi$ as follows.

**Definition 2.** *Let an equilibrium selection rule $\pi$ be a measurable function $\cup_{m=1}^M B_m \to [0,1]^M$ defined as*

$$\begin{aligned} \pi(x) &= (\pi_1(x), \pi_2(x), ..., \pi_M(x)), \\ \textit{where } \pi_m(x) &= \Pr(Y \in q_m(B_m) | X = x). \end{aligned}$$

The component $\pi_m(x)$ is the probability of selecting equilibrium $m$ conditional on $X = x$. Clearly, $\pi_m(x) = 0$ if $m \notin \Upsilon(x)$. Also, $\sum_{m=1}^{M} \pi_m(x) = 1$.

The following equation describes the relationship between the observed conditional choice probabilities $Q\,(1|1,x)$ and the equilibrium beliefs $(\sigma_1(1|1,x), \sigma_2(1|1,x))'$ at a given $x$:

$$Q\,(1|1,x) = \sum_{h \in \Upsilon(x)} \pi_h(x) \sigma^{(h)}(1|1,x), \tag{3.9}$$

where $\sigma^{(h)}(1|1,x) = (\sigma_1^{(h)}(1|1,x), \sigma_2^{(h)}(1|1,x))' = q_h(x)$. Equation (3.9) says that a conditional choice probability is a mixture of the equilibrium beliefs that are solutions to (3.7). The mixing probabilities are the corresponding components of $\pi(x)$.

### 3.3.1 The Equilibrium Behaviour and Presence of A Jump in Conditional Choice Probabilities

Multiplicity of equilibria in games with incomplete information has drawn increasing attention in the literature. Theoretically, multiple equilibria may arise when the system of equations (3.7) admits more than one solution. The main result of this section is the relationship between the equilibrium behaviour and the presence of a jump in the conditional choice probabilities $Q\,(1|1,x)$. In particular, we show that if the system of equations (3.7) admits a unique solution for all $x \in \mathcal{X}$, then the conditional choice probabilities $Q\,(1|1,x)$ must be continuous in $x$ on $\text{int}(\mathcal{X})$. On the other hand, if the system of equations (3.7) admits multiple solutions and the index of the realized equilibrium varies over $x$, then $Q\,(1|1,x)$ will have a jump at some $x$, under reasonable conditions. Note that $Q\,(1|1,x)$ is said to be continuous when both $Q_1\,(1|1,x)$ and $Q_2\,(1|1,x)$ are continuous. On the other hand, $Q\,(1|1,x)$ has a jump at some $x$ means that at least one of the two conditional choice probabilities has a jump at $x$.

We first make clear the meaning of uniqueness and multiplicity of equilibria. We classify the game into one of the following three categories.

**(C1)** The system of equations (3.7) admits a unique solution in $\sigma$ for all $x \in \mathcal{X}$.

**(C2)** There exists a set $\mathcal{X}_A \subset \mathcal{X}$ with $\Pr(X \in \mathcal{X}_A) > 0$, such that the system of equations (3.7) admits multiple solutions in $\sigma$ for all $x \in \mathcal{X}_A$.

Moreover, for any $x \in \mathcal{X}_A$, there exists an $m^*$, which may depend on $x$, such that $\pi_{m^*}(x) = 1$.

**(C3)** There exists a set $\mathcal{X}_A \subset \mathcal{X}$ with $\Pr(X \in \mathcal{X}_A) > 0$, such that the system of equations (3.7) admits multiple solutions in $\sigma$ for all $x \in \mathcal{X}_A$. Moreover, there exists a set $\mathcal{X}_B \subset \mathcal{X}_A$ with $\Pr(X \in \mathcal{X}_B) > 0$, satisfying the following conditions: for any $x \in \mathcal{X}_B$, there are $m_1 \neq m_2$ such that $\pi_{m_1}(x) > 0$ and $\pi_{m_2}(x) > 0$.

Note that this classification does not include an exceptional case in which structural equations (3.7) admit multiple solutions on a non-empty but zero-measure set of $x$. Also, it does not include another exceptional case in which the set $\{x \in \mathcal{X} : \pi_{m_1}(x) > 0, \pi_{m_2}(x) > 0, \text{ for some } m_1, m_2\}$ is non-empty but has a zero measure. We do not consider those cases in this chapter. In the literature of games with incomplete information, usually games fitting into Category (C1) or (C2) are viewed as having a unique equilibrium, whereas games in Category (C3) are viewed as having multiple equilibria.

We further classify (C2) into two sub-categories:

**(C2-1)** There is an $m^*$, which does not depend on $x$, such that $\pi_{m^*}(x) = 1$ for all $x \in \mathcal{X}$.

**(C2-2)** Two disjoint subsets of $\mathcal{X}$, $C_1$ and $C_2$, satisfying $\bar{C}_1 \cap \bar{C}_2 \neq \varnothing, \Pr(X \in C_1) > 0, \Pr(X \in C_2) > 0$, have the following property:

$$\pi_{m_1^*}(x) = 1 \text{ for } x \in C_1, \pi_{m_2^*}(x) = 1, \text{ for } x \in C_2,$$

for some $m_1^* \neq m_2^*$.

Similarly, we do not consider an exceptional case where the set $\{x \in \mathcal{X} : \pi_{m^*}(x) = 1\}$ is non-empty but has a zero measure. Sub-category (C2-1) can be viewed as in Category (C1). Indeed, the system of structural equations (3.7) coupled with the equilibrium rule $\pi_{m^*}(x) = 1$ and for an $m^*$ and for all $x \in \mathcal{X}$, is observationally equivalent to the structural equations $\sigma = q_{m^*}(x)$ and $\mathcal{X} = B_{m^*}$. Clearly, the latter fits into Category (C1). Therefore, in the rest of this chapter, we treat Category (C1) and Sub-category (C2-1) as a whole.

The following proposition associates the uniqueness of solution with continuity of the conditional choice probabilities.

**Proposition 1.** *Under Assumptions 1-3, if the game fits into Category (C1) or Subcategory (C2-1), the conditional choice probabilities* $Q(1|1, x)$ *will be twice continuously differentiable in* $x$ *on* $\text{int}(\mathcal{X})$.

As a result, when we are testing the null hypothesis that $Q(1|1,x)$ is continuous in $x$ on $\mathrm{int}(\mathcal{X})$, we are actually testing a sufficient condition for uniqueness of equilibrium.

Now let us investigate what happens when the game has multiple equilibria, (i.e., the game fits into Category (C3)). We will show that under some reasonable conditions on the structural equations (3.7), when there are multiple equilibria, the conditional choice probabilities $Q(1|1,x)$ will have a jump at certain $x$, except for some special equilibrium selection rules. Here a jump at point $x_d$ (we use the word jump and discontinuity interchangeably throughout this chapter) means that the limit of $Q(1|1,x)$ towards $x_d$ does not equal in every direction. That is, there are disjoint subsets $C_1$ and $C_2$ of $\mathcal{X}$ with $\Pr(X \in C_1) > 0, \Pr(X \in C_2) > 0$, and some $x_d \in \bar{C}_1 \cap \bar{C}_2$ such that for $k = 1$ *or* 2,

$$\left| \lim_{x \to x_d, x \in C_1} Q_k(1|1,x) - \lim_{x \to x_d, x \in C_2} Q_k(1|1,x) \right| = \delta > 0.$$

The main source for such a jump lies in the equilibrium behaviour[18]. When the structural equations (3.7) admit multiple solutions, an important and economic-relevant scenario is that the set of equilibrium indices $\Upsilon(x)$ varies over $\mathcal{X}$. (A typical case is that the number of solutions to (3.7) changes over $\mathcal{X}$). Consider what happens at the boundary between regions of $x$ with different equilibria indices sets. Suppose that there is a region of $x$ where (3.7) has a unique solution and another region where (3.7) has more than one solution. Equation (3.9) implies that at one side of the boundary, $Q(1|1,x)$ equals the unique solution, whereas on the other side of the boundary, $Q(1|1,x)$ is a mixture of more than one solution. As a result, except for some special equilibrium selection rules, $Q(1|1,x)$ will have a jump at that boundary. This idea will be formalized in Proposition 2.

Let us give conditions for a jump to happen when the game has multiple equilibria. Recall the equilibria indices set is defined in (3.8), if there is a set $C \in \mathcal{X}$ such that $\Upsilon(x)$ is the same for all $x \in C$, we say that $\Upsilon(C)$ is the equilibria indices set on the given set $C$. Clearly, the value of $\Upsilon(C)$ is given by $\Upsilon(C) = \Upsilon(x)$ for an arbitrary $x \in C$. The following assumption requires that $\Upsilon(C)$ changes over subsets of $\mathcal{X}$.

**Assumption 4H.** If the game fits into Category (C2-2) or (C3), there is

[18] There is another possible source of jump which is related to the form of equilibrium selection rule. We will discuss it later. As the latter source needs stronger conditions on the equilibrium selection rules, we view it as a secondary source of a jump and will discuss it later.

an $l \in \{1, ..., M\}$ such that for some disjoint $C_1, C_2 \subset B_l$ satisfying $\Pr(X \in C_1) > 0$, $\Pr(X \in C_2) > 0$ and $\bar{C}_1 \cap \bar{C}_2 \neq \varnothing$, we have $\Upsilon(C_1) \neq \Upsilon(C_2)$.

Assumption 4H says that equilibrium indices set changes within some $B_l$. It includes a typical case specified in Assumption 4L, provided that Assumption 5 below holds. Assumption 4L is easier to interpret than Assumption 4H.

**Assumption 4L.** If the game fits into Category (C2-2) or (C3), there is a subset $C$ of $\mathcal{X}$ with $\Pr(X \in C) > 0$, such that equations (3.7) admit a unique solution in $\sigma$ for all $x \in C$.

Assumption 4L says that the structural equations (3.7) admit a unique solution for some values of $x$ while admit multiple solutions for other values of $x$ (by the definition of (C3)). This often happens in games with multiple equilibria. Multiple equilibria in players' beliefs can be viewed as a result of a coordination mechanism. When the covariates take a very large (or small) value, the effect of covariates dominates the coordination mechanism, and it uniquely determines the equilibrium beliefs (i.e., there is only one solution to (3.7) given such an extremal $x$). In contrast, when the values of the covariates are moderate, the self-fulling mechanism tends to dominate, and produces (for example) three equilibria: a large one, a medium one and a small one. The following Assumption 5 is a technical condition under which Assumption 4L is sufficient for Assumption 4H.

**Assumption 5.** (i) For an arbitrary $x \in \mathrm{int}(\mathcal{X})$, if for any $\epsilon > 0$ there exists an $x_0$ such that $|x - x_0| < \epsilon$ and $\varphi(\sigma, x_0)$ has a unique solution in $\sigma$, then $\bigtriangledown_\sigma \varphi(\sigma, x)$ is invertible for any $\sigma$ satisfying $\varphi(\sigma, x) = 0$.

(ii) For any $m \in \{1, ..., M\}$, let $x_0$ be an arbitrary point $x_0 \in \bar{B}_m \backslash B_m$, and $\sigma^*$ satisfy $\varphi(\sigma^*, x_0) = 0$, $\sigma^* \neq \lim_{x \to x_0, x \in B_m} q_m(x) \equiv \sigma_m$. If $\sigma^*$ exists and there is no such $\sigma'$ that every component of $\sigma' - \sigma_m$ has the same sign as $\sigma^* - \sigma_m$ and $|\sigma' - \sigma_m| < |\sigma^* - \sigma_m|$, then the partial derivative $\bigtriangledown_\sigma \varphi(\sigma^*, x_0)$ is invertible.

Assumption 5 assumes that the partial derivative $\bigtriangledown_\sigma \varphi(\sigma, x)$ is invertible at some particular $(\sigma, x)$. Lemma 1 below establishes the relation between Assumption 4L and Assumption 4H.

**Lemma 1.** *Under Assumption 5, Assumption 4L is sufficient for Assumption 4H.*

In the following, whenever we state a result under Assumption 4H, it also holds if Assumption 4H is replaced by Assumption 4L and Assumption 5.

Note that the function $\Upsilon : \mathcal{X} \to$ the power set of $\{1, ..., M\}$, which maps an $x$ to an equilibria indices set, is fully determined by the structural equations (3.7). Given a function $\Upsilon$, we can view an equilibrium selection rule $\pi$ as an element in a class

$$\mathcal{C}(\Upsilon) = \left\{ \begin{array}{c} g : \cup_{m=1}^{M} B_m \to [0,1]^M : \sum_{h=1}^{M} g_h(x) = 1, \\ \text{and } g_j(x) = 0 \text{ if } j \notin \Upsilon(x), \text{ for all } x \in \cup_{m=1}^{M} B_m. \end{array} \right\},$$

where $g_j(x)$ is the $j$th component of $g(x) \in [0,1]^M$. Proposition 2 below states that under Assumptions 1-3 and 4H, multiplicity of equilibria in the incomplete information game leads to a jump in the conditional choice probabilities $Q(1|1,x)$ at some $x$, except for equilibrium selection rules that take values of a zero measure "around" $x$'s at the boundary between regions with different equilibria indices sets.

**Proposition 2.** *Under Assumptions 1-3 and 4H, if the game fits into Category (C3), the conditional choice probabilities* $Q(1|1,x)$ *will have a jump at some* $x \in \text{int}(\mathcal{X})$*, except for a class of equilibrium selection rules* $\mathcal{C}_{E_1}(\Upsilon)$*, and the set*

$$\{(\lim_{x' \in C_1, x' \to x} \pi(x'), \lim_{x' \in C_2, x' \to x} \pi(x')) : \pi \in \mathcal{C}_{E_1}(\Upsilon)\}$$

*has a zero measure in*

$$\{(\lim_{x' \in C_1, x' \to x} \pi(x'), \lim_{x' \in C_2, x' \to x} \pi(x')) : \pi \in \mathcal{C}(\Upsilon)\},$$

*for all* $x \in \bar{C}_1 \cap \bar{C}_2$*, and all disjoint* $C_1, C_2 \subset B_l$*, for some* $l \in \{1, ..., M\}$ *such that* $\Pr(X \in C_1) > 0$, $\Pr(X \in C_2) > 0$, $\bar{C}_1 \cap \bar{C}_2 \neq \varnothing$, $\Upsilon(C_1) \neq \Upsilon(C_2)$.

The existence of such $x$'s in Proposition 2 is guaranteed by Assumption 4H (or Assumption 4L and Assumption 5). Consider the numerical example in the Section 3.3.2. Its normal-form is in Table 3.2. For some values of parameters, the structural equations (3.11) has either one or three solutions depending on the $x$ value. It satisfies Assumption 4L, and there are $C_1, C_2$ satisfying Assumption 4H with $l = 1$, $\Upsilon(C_1) = \{1\}$ and $\Upsilon(C_2) = \{1, 2, 3\}$. For an arbitrary $x_d \in \bar{C}_1 \cap \bar{C}_2$, equation (3.9) implies that

$$\begin{aligned}
\lim_{x\to x_d, x\in C_1} Q(1|1,x) &= \sigma^{(1)}(1|1,x_d), \\
\lim_{x\to x_d, x\in C_2} Q(1|1,x) &= \lim_{x\to x_d, x\in C_2} \pi_1(x)\sigma^{(1)}(1|1,x_d) \\
&\quad + \lim_{x\to x_d, x\in C_2} \pi_2(x) \lim_{x\to x_d, x\in C_2} \sigma^{(2)}(1|1,x) \\
&\quad + \lim_{x\to x_d, x\in C_2} \pi_3(x) \lim_{x\to x_d, x\in C_2} \sigma^{(3)}(1|1,x).
\end{aligned}$$

Hence

$$\begin{aligned}
& \lim_{x\to x_d, x\in C_1} Q(1|1,x) - \lim_{x\to x_d, x\in C_2} Q(1|1,x) \\
= & \left(1 - \lim_{x\to x_d, x\in C_2} \pi_1(x)\right)\sigma^{(1)}(1|1,x_d) \\
& - \lim_{x\to x_d, x\in C_2} \pi_2(x) \lim_{x\to x_d, x\in C_2} \sigma^{(2)}(1|1,x) \\
& - \lim_{x\to x_d, x\in C_2} \pi_3(x) \lim_{x\to x_d, x\in C_2} \sigma^{(3)}(1|1,x).
\end{aligned}$$

Consider a class of equilibrium selection rules in which $(\pi_1, \pi_2, 1-\pi_1-\pi_2)$ are the probabilities of picking up each equilibrium when the structural equations have three solutions. Now, a desired jump will arise at $x_d$ as long as

$$\begin{aligned}
& (1-\pi_1)\sigma^{(1)}(1|1,x_d) - \pi_2 \lim_{x\to x_d, x\in C_2} \sigma^{(2)}(1|1,x) \\
& - (1-\pi_1-\pi_2) \lim_{x\to x_d, x\in C_2} \sigma^{(3)}(1|1,x) \\
\neq & \; 0.
\end{aligned}$$

Rearranging the terms leads to

$$\begin{aligned}
& \left(\lim_{x\to x_d, x\in C_2} \sigma^{(3)}(1|1,x) - \sigma^{(1)}(1|1,x_d)\right)\pi_1 \\
& + \left(\lim_{x\to x_d, x\in C_2} \sigma^{(3)}(1|1,x) - \lim_{x\to x_d, x\in C_2} \sigma^{(2)}(1|1,x)\right)\pi_2 \\
& + \sigma^{(1)}(1|1,x_d) - \lim_{x\to x_d, x\in C_2} \sigma^{(3)}(1|1,x) \neq 0,
\end{aligned}$$

Note that $\sigma^{(1)}(1|1,x_d) \neq \lim_{x\to x_d, x\in C_2} \sigma^{(3)}(1|1,x)$ for all $x_d \in \bar{C}_1 \cap \bar{C}_2$. [19] As a result, the conditional choice probabilities $Q(1|1,x)$ will have a jump,

[19] Suppose to the contrary, $\sigma^{(1)}(1|1,x_d) = \lim_{x\to x_d, x\in C_2} \sigma^{(3)}(1|1,x)$ for some $x_d \in \bar{C}_1 \cap$

except for some artificially chosen $(\pi_1, \pi_2) \in \{(\pi_1, \pi_2) \in [0,1]^2 : \pi_1 + \pi_2 \leq 1\}$ satisfying the following restriction:

$$\begin{aligned} \alpha\pi_1 + \beta\pi_2 &= \gamma, \\ \text{where } \alpha &= \lim_{x \to x_d, x \in C_2} \sigma^{(3)}(1|1,x) - \sigma^{(1)}(1|1,x_d) \neq 0, \\ \beta &= \lim_{x \to x_d, x \in C_2} \sigma^{(3)}(1|1,x) - \lim_{x \to x_d, x \in C_2} \sigma^{(2)}(1|1,x), \\ \gamma &= \sigma^{(1)}(1|1,x_d) - \lim_{x \to x_d, x \in C_2} \sigma^{(3)}(1|1,x), \\ \text{for all } x_d &\in \bar{C}_1 \cap \bar{C}_2, \text{ and all } C_1, C_2 \text{ satisfying 4H with } l = 1, \end{aligned} \tag{3.10}$$

Clearly the set $\{(\pi_1, \pi_2) \in [0,1]^2 : \pi_1 + \pi_2 \leq 1, \text{and (3.10) holds}\}$ has a zero measure in $\{(\pi_1, \pi_2) \in [0,1]^2 : \pi_1 + \pi_2 \leq 1\}$. Note that here the equilibrium selection rule does not depend on $x$ within the regions where structural equations (3.7) have three solutions. As a result, the above "zero measure" argument for the equilibrium selection rule holds for any $x$ that leads to three solutions.

The main idea of Proposition 2 is that the conditional choice probabilities $Q(1|1,x)$ will have a jump at the boundary between regions of $x$ with distinct equilibrium indices sets (especially at the boundary between regions have different numbers of solutions to the structural equations (3.7)), except for some special equilibrium selection rules. However, such boundaries are not the only type of locations jumps may happen. A jump can also occur in the interior of a region with the same equilibria indices set, if the equilibrium selection rule itself has a jump in $x$. Definition 3 characterizes such equilibrium selection rules.

**Definition 3.** *We say an equilibrium selection rule $\pi$ has a jump within an equilibrium indices set $\Upsilon(D)$ if there is a subset $D \subset \text{int}(\mathcal{X})$ with $\Upsilon(D)$ defined, and the following condition is satisfied: there are two disjoint subsets $D_1, D_2 \subset D$, $\Pr(X \in D_1) > 0, \Pr(X \in D_2) > 0$, $\bar{D}_1 \cap \bar{D}_2 \neq \varnothing$, such that for all $x \in \bar{D}_1 \cap \bar{D}_2$, we have*

$$\left| \lim_{x' \to x, x' \in D_1} \pi_{h^*}(x') - \lim_{x' \to x, x' \in D_2} \pi_{h^*}(x') \right| = \delta > 0, \textit{ for some } h^* \in \Upsilon(D).$$

$\bar{C}_2$. This means $\lim_{x \to x_d, x \in C_1} q_1(x) = \lim_{x \to x_d, x \in C_2} q_3(x)$ for some $x_d \in \bar{C}_1 \cap \bar{C}_2$. Note that $x_d \in B_1$ and $x_d \in \bar{B}_3$, we have $q_1(x') = q_3(x')$ for some $x' \in B(x_d, r) \cap B_1 \cap B_3$, where $B(x_d, r)$ is a ball centred in $x_d$ and with radius $r$. $B(x_d, r) \cap B_1 \cap B_3 \neq \varnothing$ because $x_d \in \bar{C}_1 \cap \bar{C}_2, C_1, C_2 \subset B_1$ (since $l = 1$ in Assumption 4L) and $C_2 \subset B_3$. This contradicts with Remark 1 below Definition 1.

*Denote as $\mathcal{C}_F(\Upsilon)$ the class of equilibrium selection rules that have a jump within an equilibria indices set.*

The next result states that if the game has multiple equilibria and the equilibrium selection rule has the property specified in Definition 3, the conditional choice probabilities $Q\,(1|1,x)$ will have a jump at some $x$, except for some special equilibria selection rules.

**Corollary 1.** *Under Assumptions 1-3, if the game fits into Category (C3) and the equilibrium selection rule $\lambda \in \mathcal{C}_F(\Upsilon)$ (i.e., the equilibrium selection rule has the property in Definition 3), the conditional choice probabilities $Q\,(1|1,x)$ will have a jump at some $x \in \mathrm{int}(\mathcal{X})$, except for a class of equilibrium selection rules $\mathcal{C}_{E_2}(\Upsilon)$, and the set*

$$\{(\lim_{x' \in D_1, x' \to x} \pi(x'), \lim_{x' \in D_2, x' \to x} \pi(x')) : \pi \in \mathcal{C}_{E_2}(\Upsilon)\}$$

*has a zero measure in*

$$\{(\lim_{x' \in D_1, x' \to x} \pi(x'), \lim_{x' \in D_2, x' \to x} \pi(x')) : \pi \in \mathcal{C}_F(\Upsilon)\},$$

*for all $x \in \bar{D}_1 \cap \bar{D}_2$ in Definition 3.*

So far we have studied the game fitting into Category (C1), Sub-category (C2-1) or Category (C3). What remains is Sub-category (C2-2). Unfortunately, the game in (C2-2) exhibits undesirable properties in the conditional choice probabilities. The following Corollary 2 shows that the game in (C2-2), which is usually regarded as having a unique equilibrium, leads to jump(s) in the conditional choice probabilities. This is because the single equilibrium present in the data jumps from one index to the other at some $x$. This is the main limitation of the discontinuity criterion in testing for multiple equilibria in discrete games with incomplete information.

**Corollary 2.** *Under Assumptions 1-3, if the game fits into Sub-category (C2-2), the conditional choice probabilities $Q\,(1|1,x)$ will have a jump at some $x \in \mathrm{int}(\mathcal{X})$.*

In sum, we have shown that under Assumptions 1-3 and 4H, the equilibrium behaviour of incomplete information games produces testable implications on continuity or discontinuity of the conditional choice probabilities $Q\,(1|1,x)$. Such relations are robust to correlated private information between players.

Table 3.2: Normal-form of the numerical example

| | $y_2 = 0$ | $y_2 = 1$ |
|---|---|---|
| $y_1 = 0$ | $0, 0$ | $0, \alpha + \beta x - u_2$ |
| $y_1 = 1$ | $\alpha + \beta x - u_1, 0$ | $\alpha + (\beta + \gamma)x - u_1, \alpha + (\beta + \gamma)x - u_2$ |

### 3.3.2 A Numerical Example

We present a numerical example to illustrate the relationship between the equilibrium behaviour and the presence of a jump in the conditional choice probabilities. Suppose that the game has a normal-form in Table 3.2.

In this example, scalar $X$ is the only covariate. Assume the joint distribution of $(U_1, U_2)'$ is

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \sim N\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right).$$

The resulting equilibrium characterizing equations are:

$$\begin{bmatrix} \sigma_1 \\ \sigma_2 \end{bmatrix} = \begin{bmatrix} G_{1,2}(\alpha + \beta x + \gamma\sigma_2, \alpha + \beta x + \gamma\sigma_1)/G_2(\alpha + \beta x + \gamma\sigma_1) \\ G_{1,2}(\alpha + \beta x + \gamma\sigma_2, \alpha + \beta x + \gamma\sigma_1)/G_1(\alpha + \beta x + \gamma\sigma_2) \end{bmatrix}, \tag{3.11}$$

where $G_{1,2}$ is the bivariate CDF of $(U_1, U_2)'$, $G_k$ is the marginal CDF of $U_k$, and $\sigma_k$ is player $-k$'s belief that player $k$ chooses action 1, conditional on player $-k$ choosing action 1.

Consider two games with different values of parameters $(\alpha, \beta, \gamma, \rho)'$ specified in Table 3.3. Game 1 fits into Category (C1) while Game 2 fits into Category (C3). Their parameters only differ in $\gamma$. When $\gamma > 0$, the game with normal-form in (3.2) is a coordination game in which $\gamma$ determines the strength of interaction. For Game 1, the structural equations (3.11) always have a unique solution, while for Game 2, the structural equations (3.11) have one or three solutions depending on the value of $x$. Figure 3.1 plots the equilibrium behaviour of Game 2 at $x = 0.4$ (upper) and $x = 0.5$ (lower). In Figure 3.1, the solid line is the best response function of belief $\sigma_1$ in terms of $\sigma_2$ at a fixed $x$. Because of the symmetry of the game, the equilibrium belief $\sigma_1$ (and $\sigma_2$) is the intersection of the best response function and the 45 degree line (the dashed line). At $x = 0.4$, the only equilibrium belief is $\sigma_1 = \sigma_2 = 0.456$, while at $x = 0.5$, there are three equilibrium beliefs: 0.460, 0.698 and 0.916.

Figure 3.1: The equilibrium belief(s) $\sigma_1$
$\alpha = 0.25, \beta = -3.5, \gamma = 7, \rho = 0.2$

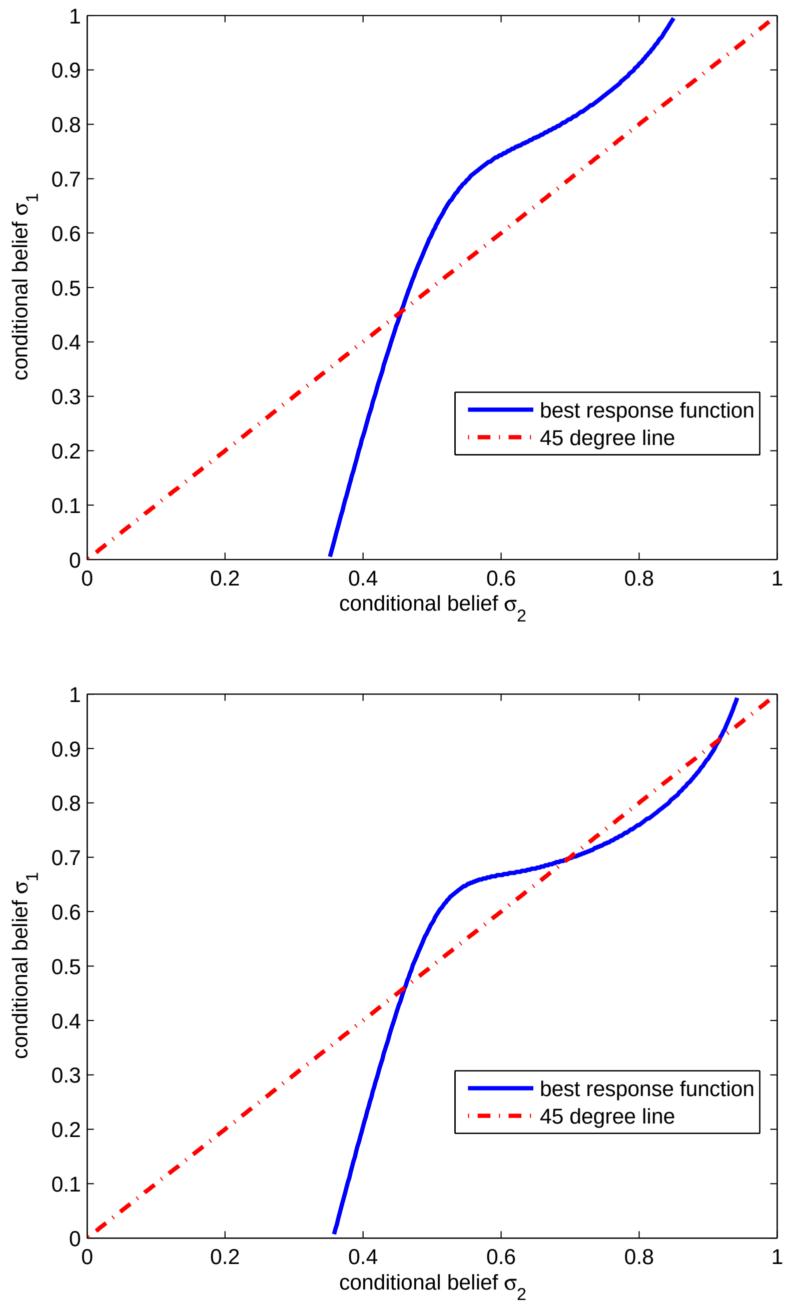


Table 3.3: Parameters for the games

| | Game 1 | Game 2 |
|---|---|---|
| $(\alpha, \beta, \gamma, \rho)'$ | $(0.25, -3.5, 3.5, 0.2)'$ | $(0.25, -3.5, 7, 0.2)'$ |
| Support of $X$ | $[0.35, 0.55]$ | $[0.35, 0.55]$ |
| # of solutions[20] | 1 | 1 or 3 |

Figure 3.2 plots the equilibria correspondence from the covariate value $x$ to the equilibrium belief(s) $\sigma_1$, for Game 2. When $x$ is smaller than 0.464, the structural equations (3.11) admit a unique solution (the solid line). On the other hand, when $x$ is larger than 0.464, the structural equations (3.11) admit three solutions (the solid, dash and dot-dash line). The functions $q_1$, $q_2$ and $q_3$ are three branches of the correspondence.

We examine the observed conditional choice probability $Q_1(1|1, x) = \Pr(Y_1 = 1|Y_2 = 1, X = x)$. Figure 3.3 plots $Q_1(1|1, x)$ as a function of $x$, for Game 1 (upper) and Game 2 (lower).

For Game 1, the conditional choice probability $Q_1(1|1, x)$ is continuous on the support of $X$. For Game 2, if the equilibrium selection rule assigns equal probabilities to three solutions whenever they are available, $Q_1(1|1, x)$ has a jump at $x = 0.464$.

### 3.3.3 An Extension: Unobserved Heterogeneity

In this section, we allow for heterogeneity in the payoff functions. Here the heterogeneity refers to various types of game environments determined by a random variable $T$, before the players make choices. The payoff function for player $k$ can be written as

$$\pi_k(Y_k, Y_{-k}, X_k, T) - U_k(Y_k). \tag{3.12}$$

As in (3.1), we make normalizations $U_k(0) = 0$ and $U_k(1) = U_k$. The departure from (3.1) is that the function $\pi_k$ now is $t$-dependent. The normal-form of the game is in Table 3.12, given $t \in \{1, ..., \bar{t}\}$. In each observation $i$, $(Y_{1i}, Y_{2i}, X'_{1i}, X'_{2i})$ are observed by the econometrician, $T_i$ is observed by players but not by the econometrician. $U_{ki}$ is the private information to the player $k$, $k = 1, 2$. We assume that $T_i$ can take values in $\{1, ..., \bar{t}\}$ and with probability $\Pr(T_i = t) = \lambda_t$, for $t = 1, ..., \bar{t}$. However, the discrete support of $T$ is not essential. $T$ can be generalized to a continuous random

Figure 3.2: Equilibria correspondence from covariate $x$ to belief $\sigma_1$
$\alpha = 0.25, \beta = -3.5, \gamma = 7, \rho = 0.2$

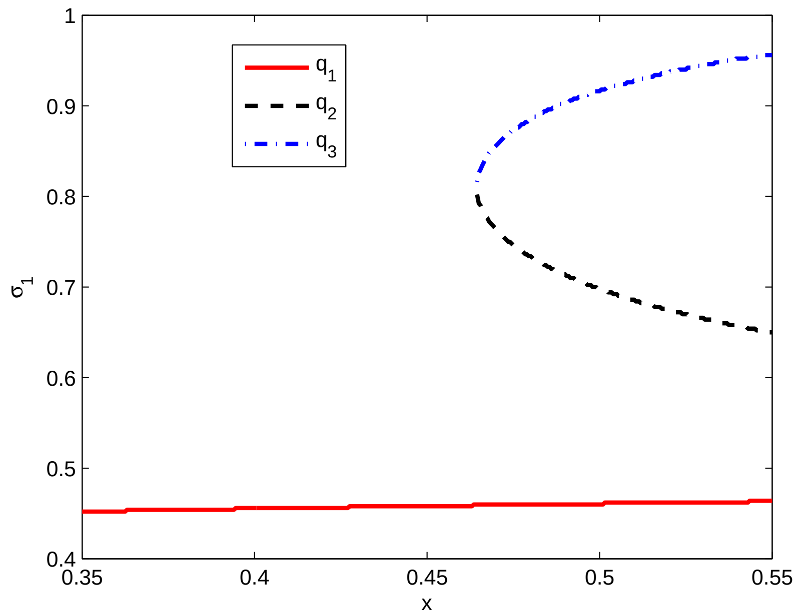
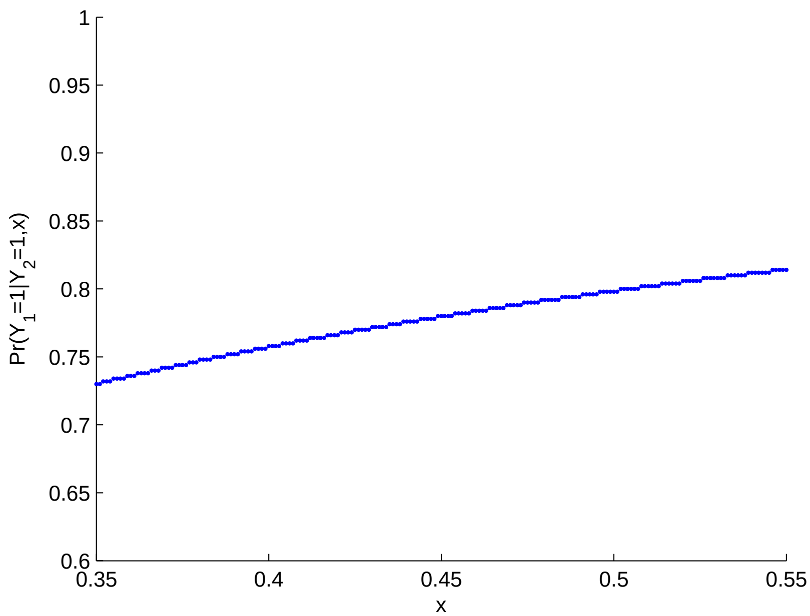


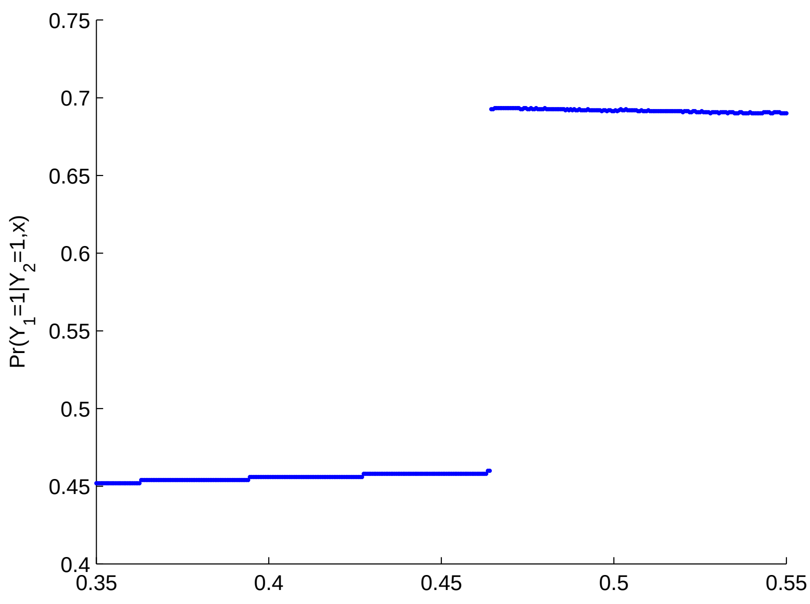

Table 3.4: Normal-form of the game with $T = t$

| | $y_2 = 0$ | $y_2 = 1$ |
|---|---|---|
| $y_1 = 0$ | $0, 0$ | $0, \pi_2(0, 1, x_2, t) - u_2$ |
| $y_1 = 1$ | $\pi_1(1, 0, x_1, t) - u_1, 0$ | $\pi_1(1, 1, x_2, t) - u_1, \pi_2(1, 1, x_2, t) - u_2$ |

Figure 3.3: Observed conditional choice probability
$\alpha = 0.25, \beta = -3.5, \gamma = 3.5, \rho = 0.2$




$\alpha = 0.25, \beta = -3.5, \gamma = 7, \rho = 0.2$, equilibrium selection rule $\pi = (1/3, 1/3, 1/3)$



variable, though we do not cover it in this chapter. The real restriction on the heterogeneity $T$ is the part (iv) of the following Assumption 2'.

**Assumption 2'.** (i) For each realization $t$ of $T$, the random vector $U = (U_1, U_2) \in \mathbf{R}^2$ is continuously distributed with a conditional CDF $F_{1,2}^t(\cdot)$, which is twice continuously differentiable.

(ii) The unobservable $U$ is independent of $X$ conditional on $T$.

(iii) Conditional on $T = t \in \{1, ..., \bar{t}\}$, the support of $(U_1, U_2)$ satisfies

$$\begin{aligned} \inf \mathcal{U}_1 &< \inf_{x \in \mathcal{X}} \min \left(\pi_1(1, 0, x_1, t), \pi_1(1, 0, x_1, t) + \Delta_1(x_1, t)\right), \\ \inf \mathcal{U}_2 &< \inf_{x \in \mathcal{X}} \min \left(\pi_2(1, 0, x_2, t), \pi_2^t(1, 0, x_2, t) + \Delta_2(x_2, t)\right), \end{aligned}$$

where $\Delta_k(x_k, t) = \pi_k^t(1, 1, x_k, t) - \pi_k^t(1, 0, x_k, t)$ for $k = 1, 2$.

(iv) $T$ is independent of $X$.

Assumption 2'(i)-(iii) are modified versions of Assumption 2. Assumption 2' (iv) says that the unobserved heterogeneity is independent of covariates $X$. An implication is that the support $\mathcal{X}$ is the same for all $t$. Otherwise, the heterogeneity depending on covariates by itself may create jump(s) in the conditional choice probabilities, which contaminates the relationship between the equilibrium behaviour and the presence of jump.

Fixing $T = t$, the corresponding game is the same as the one we have analysed before. The equilibrium characterizing equations become

$$\begin{bmatrix} \sigma_1^t(1|1, x) \\ \sigma_2^t(1|1, x) \end{bmatrix} = \begin{bmatrix} \frac{F_{1,2}^t\left(\pi_1(1,0,x_1,t)+\sigma_2^t(1|1,x)\Delta_1(x_1,t), \pi_2(1,0,x_2,t)+\sigma_1^t(1|1,x)\Delta_2(x_2,t)\right)}{F_2^t\left(\pi_2(1,0,x_2,t)+\sigma_1^t(1|1,x)\Delta_2(x_2,t)\right)} \\ \frac{F_{1,2}^t\left(\pi_1(1,0,x_1,t)+\sigma_2^t(1|1,x)\Delta_1(1,1,x_1), \pi_2(1,0,x_2,t)+\sigma_1^t(1|1,x)\Delta_2(x_2,t)\right)}{F_1^t\left(\pi_1(1,0,x_1,t)+\sigma_2^t(1|1,x)\Delta_1(x_1,t)\right)} \end{bmatrix}, \tag{3.13}$$

for $t \in \{1, ..., \bar{t}\}$, where $(\sigma_1^t(1|1, x), \sigma_2^t(1|1, x))'$ is the pair of equilibrium beliefs for the game with $T = t$. For each $t$, denote equation (3.13) as

$$\varphi^t(\sigma, x) = 0. \tag{3.14}$$

The following Definition 1', 2' and Assumption 3', 4H' are modified versions of their counterparts in the previous subsection. Thus we state them without further discussions.

**Definition 1'.** *For each $t \in \{1, ..., \bar{t}\}$, let $M^t$ be the smallest number such that there exists a family of twice continuously differentiable functions*

$$\left\{q_1^t(\cdot), q_2^t(\cdot), ..., q_{M^t}^t(\cdot)\right\},$$

*where* $q_m^t : B_m^t \to [0,1]^2$, $B_m^t$ *is an open and connected subset of* $\mathcal{X}$, *and the following conditions are satisfied*:

*(i) For any* $(\sigma, x) \in [0,1]^2 \times B_m^t$ *satisfying* $\varphi^t(\sigma, x) = 0$, *we have* $\sigma = q_m^t(x)$. *Furthermore,* $\varphi^t(q_m^t(x), x) = 0$ *for all* $x \in B_m^t$.

*(ii)* $\cup_{m=1}^M \bar{B}_m^t = \mathcal{X}$. *(Here* $\bar{B}_m^t$ *denotes the closure of* $B_m^t$.*)*

Like Remark 1, without loss of generality, we can further assume that for any $x \in B_m^t \cap B_k^t$ and $m \neq k$, $q_m^t(x) \neq q_k^t(x)$.

**Assumption 3'.** For each $t \in \{1, ..., \bar{t}\}$,
(i) the system of equations (3.14) has at least one solution in $\sigma$;
(ii) the constant $M^t$ in Definition 1 is finite;
(iii) if the system of equations (3.14) admits a unique solution in $\sigma$ for all $x \in \mathcal{X}$, the partial derivative $\triangledown_\sigma \varphi^t(\sigma, x)$ is invertible for all $(\sigma, x) \in (0,1)^2 \times \text{int}(\mathcal{X})$.

If $\sigma \in q_m^t(B_m^t)$, call it equilibrium $m$ under $T = t$ and denote it as $\sigma^{t(m)}(1|1, x)$. If $\sigma \in q_m^t(B_m^t)$ for multiple $m$'s, we take the smallest one as the index. Similar to (3.8), we define $\Upsilon^t(x)$ as the equilibria indices set for a generic value $x$, conditional on $T = t$,

$$\Upsilon^t(x) \equiv \{m : \varphi^t\left(q_m^t(x), x\right) = 0\}.$$

For $T = t$, we define the equilibrium selection rule $\pi^t$.

**Definition 2'.** *For every* $t \in \{1, ..., t\}$, *let an equilibrium selection rule* $\pi^t$ *be a measurable function* $\cup_{m=1}^M B_m^t \to [0,1]^M$ *defined as*

$$\begin{aligned} \pi^t(x) &= (\pi_1^t(x), \pi_2^t(x), ..., \pi_M^t(x)), \\ \textit{where } \pi_m^t(x) &= \Pr\left(Y \in q_m^t(B_m^t) | X = x\right). \end{aligned}$$

In the presence of unobserved heterogeneity $T$, the observed conditional choice probabilities can be written as follows

$$Q\,(1|1,x) = \sum_{t=1}^{\bar{t}} \lambda_t Q^t\,(1|1,x) = \sum_{t=1}^{\bar{t}} \lambda_t \sum_{h \in \Upsilon^t(x)} \pi_h^t(x) \sigma^{t(h)}(1|1,x), \qquad (3.15)$$

where $\sigma^{t(h)}(1|1, x) = q_h^t(x)$, as defined in Definition 1'. We can see that the heterogeneity adds another layer of mixture to the expression of conditional choice probabilities. Then we state a modified version of Assumption 4H.

**Assumption 4H'.** For any $t \in \{1, ..., \bar{t}\}$, if the game fits into Category (C2-2) or (C3), there is an $l^t \in \{1, ..., M^t\}$ such that for some disjoint $C_1^t, C_2^t \subset B_l$ satisfying $\Pr(X \in C_1^t) > 0$, $\Pr(X \in C_2^t) > 0$ and $\bar{C}_1^t \cap \bar{C}_2^t \neq \varnothing$, we have $\Upsilon^t(C_1^t) \neq \Upsilon^t(C_2^t)$.

The following proposition establishes the relation from the equilibrium behaviour to the presence of jump(s) in the conditional choice probabilities, for incomplete information games with unobserved heterogeneity $T$. The main idea is that for any fixed $t$, the relation is the same as the previous subsection. In addition, such a relation remains after the linear combination of $t$-specific components, if the jumps in

$$Q^t(1|1,x) = \sum_{h \in \Upsilon^t(x)} \pi_h^t(x) \sigma^{t(h)}(1|1,x)$$

do not cancel out across $t$ at some $x$.

**Proposition 3.** *(i) Suppose Assumptions 1'-3' hold, if for all $t \in \{1, ..., \bar{t}\}$, the game fits into Category (C1) or Sub-category (C2-1), the conditional choice probabilities $Q(1|1,x)$ will be twice continuously differentiable in $x$ for all $x \in int(\mathcal{X})$.*

*(ii) Suppose Assumptions 1'-3' hold, and there exists a nonempty set $\mathcal{T}_D \subset \{1, ..., \bar{t}\}$ such that for all $t \in \mathcal{T}_D$ the game fits into Category (C3). For each $t \in \{1, ..., \bar{t}\}$, let $\mathcal{X}_D^t = \{x \in int(\mathcal{X}) : Q^t(1|1,x) \text{ has a jump at } x\}$. If for some $t \in \{1, ..., \bar{t}\}$, $\mathcal{X}_D^t \neq \varnothing$, and the following condition COND is satisfied, the conditional choice probabilities $Q(1|1,x)$ will have a jump at some $x \in \mathrm{int}(\mathcal{X})$.*

*COND. There is a $t$ and some $x_d \in \mathcal{X}_D^t$ such that*

$$\sum_{t \in \{t' : x_d \in \mathcal{X}_D^{t'}\}} \lambda_t Q^t(1|1,x)$$

*has a jump at $x = x_d$.*

*(iii) Suppose Assumptions 1'-3' hold, and there exists a nonempty set $\mathcal{T}_D \subset \{1, ..., \bar{t}\}$ such that for all $t \in \mathcal{T}_D$ the game fits into Sub-category (C2-2). Define $\mathcal{T}_D$ and $\mathcal{X}_D^t$* as *in part (ii). Then the conditional choice predictabilities $Q(1|1,x)$ will have a jump at some $x \in \mathrm{int}(\mathcal{X})$, if the condition COND in (ii) is satisfied.*

Consider Proposition 3(ii), for any $t \in \mathcal{T}_D$, if Assumption 4H' holds, we can use Proposition 2 to obtain $\mathcal{X}_D^t \neq \varnothing$, except for some special equilibrium selection rules. This can also be achieved by Corollary 1, provided that the equilibrium selection rule $\pi^t$ satisfies the property in Definition 3.

In sum, we have established the relationship between the equilibrium behaviour and the presence of jump(s) in the conditional choice probabilities when the payoff functions of the game contain heterogeneity unobserved by the econometrician and the heterogeneity is independent of the covariates. The observed conditional choice probabilities are mixtures of the choice probabilities given a fixed value of the heterogeneity. Hence, the relation from equilibrium behaviour to the jump(s) remains in the presence of unobserved heterogeneity.

### 3.3.4 Testing for Multiple Equilibria via Discontinuity

Proposition 1 to 3 show that in incomplete information games, the equilibrium behaviour of the game produces testable implications on continuity or discontinuity of the conditional choice probabilities. As a result, continuity or discontinuity of the conditional probabilities provides information about the equilibrium behaviour. Under Assumptions 1-3 and 4H, we can transform the problem of testing

$$\begin{aligned} H_0 &: \text{ The game fits into Category (C1) or Subcategory (C2-1)} \\ H_1 &: \text{ The game fits into Category (C3) or Subcategory (C2-2).} \end{aligned}$$

into testing

$$\begin{aligned} H_0' &: \ Q_k\left(1|1,x\right) \text{ is continuous in } x \text{ for all } x \in \operatorname{int}(\mathcal{X}) \text{ for all } k=1,2. \\ H_1' &: \ Q_k\left(1|1,x\right) \text{ has jump(s) at some } x \in \operatorname{int}(\mathcal{X}) \text{ and some } k \in \{1,2\}. \end{aligned}$$

Under Assumptions 1-3 and 4H, if $H_0'$ is not rejected, the data provides no evidence against the null hypothesis that the system of structural equations has an unique solution for all $x$. The union of Category (C1) and (C2-1) is usually regarded as a subset of games with a unique equilibrium. However, sometimes the econometrician did impose an assumption that the game fits into (C1) or (C2-1). For instance, Aradillas-Lopez (2010) proposed a semi-parametric estimation strategies for incomplete information games with parametric payoff functions, under a “unique equilibrium” assumption that corresponds to the union of (C1) and (C2-1).[21] In this scenario, testing for the presence of a jump in $Q(1|1,x)$ is checking the maintained assumption in Aradillas-Lopez (2010).

Note that *none* of propositions and corollaries in this chapter requires the independence between players' private information. In other words, the implications of the equilibrium behaviour for continuity or discontinuity of the conditional choice probability are robust to correlated private information. As a comparison, tests for multiple equilibria based on conditional dependence of players' actions (see Paula and Tang (2012), Aguirregabiria and Mira (2013), among others) required the private information to be independent across players (see Assumption 2S below). This is an advantage of the discontinuity criterion.

**Assumption 2S** (Independent private information) The private information $U_1$ is independent of the private information $U_2$.

We are aware that under Assumptions 1-3 and 4H, information provided by continuity or discontinuity of the conditional choice probabilities does not fully separate the games with a unique equilibrium from those with multiple equilibria. However, it at least offers partial information about the equilibrium behaviour. The uniqueness or multiplicity of equilibria in incomplete information games is important for identification and estimation, especially when the private information is correlated among players. To the best of my knowledge, related literature such as Aradillas-Lopez (2010), Wan and Xu (2014) and Liu, Vuong, and Xu (2013) all maintained that the game has a unique equilibrium.

Moreover, knowing the presence of jump(s) in the conditional choice probabilities is useful if the econometrician wants to implement Paula and Tang (2012)'s test for multiple equilibria. Under Assumption 2S, they proposed to test the following independence restriction:

$$\Pr(Y_1 = 1, Y_2 = 1|X = x) = \Pr(Y_1 = 1|X = x)\Pr(Y_2 = 1|X = x).$$

Implementation of such a test requires the estimation of three conditional choice probabilities. If any of them has a jump at some $x$, which is possible as we have shown, the econometrician cannot directly apply the standard nonparametric estimation technique such as kernel or series estimators.

[21] Aradillas-Lopez (2010) treated "a unique equilibrium" as "a unique solution" to the (3.7) (he considered a linear payoff function). See Proposition 2, 3 and Assumption A3 (ii) in his chapter. Recall that a DGP in (C2-1) can be viewed as in (C1), if we let $\sigma = q_{m^*}(x)$ be the equilibrium characterizing equilibrium, where $m^*$ is the index of the equilibrium that always appears in the data, by the definition of (C2-1).

## 3.4 An Estimation Problem Due to Jumps

From the last section, we know that even when the payoff functions and the latent distribution are all smooth, the conditional choice probabilities may have a jump. It raises a problem for estimating the conditional choice probabilities, when the econometrician in priori does not know the presence of such jump(s). The estimation of the conditional choice probabilities is usually the first stage for parameter estimation and inference in games with incomplete information (see for example, Bajari, Hong, Krainer and Nekipelov (2010), Paula and Tang (2012), Wan and Xu (2014), Aguirregabiria and Mira (2013), among others). If the conditional choice probabilities have a jump, the standard nonparametric estimation technique such as kernel or series estimators do not apply directly. On the other hand, because the conditional probability may not have a jump, the usual jump location estimator does not apply either (to be discussed in this section).

This section reviews a jump location estimator and explains why it does not apply immediately when the presence of a jump in the conditional choice probabilities is unknown. To simplify arguments, we assume Assumption 2S to hold in this section. The conclusion can be generalized to cases without Assumption 2S, as we can use the sub-sample with $Y_{-k} = 1$.

Under Assumption 2S, the choice of the player $-k$ does not appear in the conditioning set of the choice probability of the player $k$, that is,

$$Q_k\left(1|1,x\right) = Q_k\left(1|x\right) \equiv \Pr\left(Y_k = 1|X = x\right), \text{ for } k = 1, 2.$$

Under Assumption 2S, Aguirregabiria and Mira (2013) noticed that the conditional probabilities $\Pr\left(Y_i = (1,1)'|X_i = x\right)$ may be discontinuous at some $x$ and "the econometrician does not know, ex-ante, the number and the location of these discontinuity points, and this complicates the application of smooth nonparametric estimators...". The solution they proposed is to first use the method developed by Müller (1992) and Delgado and Hidalgo (2000) to estimate the location(s) of the jump(s). However, their jump location estimator relies on the existence of a jump in the regression function, or in other words, the largest possible jump size must be strictly positive. To see this, note that the jump size appears multiplicatively in the asymptotic covariance of the jump location estimator (see equation (3.8) in Müller (1992) and Theorem 2 in Delgado and Hidalgo (2000)). Here we briefly review the jump location estimator proposed by Müller (1992). Müller (1992) considered the following nonparametric regression model for a univariate $X$.

$$Y = g(X) + U,$$

where $Y \in \mathbf{R}, X \in \mathbf{R}, U \in \mathbf{R}, \mathbf{E}[U|X] = 0, \mathbf{E}[U^2|X] = \sigma^2$. The departure from the standard univariate regression model is that $g(x)$ has a jump at $x = \tau$. In particular,

$$g(x) = f(x) + \Delta\{x > \tau\},$$

where function $f(x)$ is twice continuously differentiable. The unknown location of jump $\tau$ is the object of interest and $\Delta$ measures the jump size. Müller (1992) assumed that $\Delta > 0$ and introduce the local jump as

$$\Delta(x) = g^+(x) - g^-(x),$$

where

$$g^+(x) = \lim g_{t\downarrow 0}(x+t),\ g^-(x) = \lim g_{t\downarrow 0}(x-t).$$

In Müller (1992), $\tau$ is estimated by

$$\begin{aligned} \hat{\tau} &= \arg\max_x \hat{\Delta}(x), \\ \hat{\Delta}(x) &= \hat{g}^+(x) - \hat{g}^-(x), \end{aligned}$$

where $\hat{g}^+(x)$ and $\hat{g}^-(x)$ are constructed using one-sided kernels. The asymptotic distribution of $\hat{\tau}$ is obtained from the weak convergence of a local deviation process. Let $\hat{\delta}(y) = \hat{\Delta}(\tau + yh)$ and define a process as the $(nh)^{-1/2}zh$ deviation of $\hat{\Delta}(\tau)$ ($n$ denotes the sample size and $h$ denotes the bandwidth),

$$\xi_n(z) = nh\left[\hat{\delta}((nh)^{-1/2}z) - \hat{\delta}(0)\right].$$

Theorem 3.1 of Müller (1992) showed that $\xi_n(z)$ weakly converges to $\xi(z)$, where

$$\begin{aligned} \xi(z) &= -\Delta z^2 K_-(0)/2 + Wz, \\ W &\sim N\left(0, 2\sigma^2 \int K_-^2(v)dv\right), \end{aligned}$$

and $K_-(\cdot)$ is the one-sided kernel function. By the construction of $\hat{\tau}$ and $\xi(z)$, we have

$$\begin{aligned} \hat{\tau} &= \tau + (nh)^{-1/2}hZ_n, \\ Z_n &= \arg\max_z \xi_n(z). \end{aligned}$$

Let $Z^* = \arg\max_z \xi(z)$, we can compute $Z^* = W/(\Delta K_-(0))$, which gives the identification of $Z^*$, if $\Delta > 0$. Corollary 3.1 of Müller (1992) established the asymptotic distribution of estimator $\hat{\tau}$ as follows,

$$(nh)^{1/2}(\hat{\tau} - \tau) \to_d N\left(0, \frac{2\sigma^2}{(\Delta K_-(0))^2}\int K_-^2(v)dv\right). \tag{3.16}$$

Given that, the asymptotic distribution of the estimated jump size, $\hat{\Delta}(\hat{\tau})$, can be derived as (see Corollary 3.2 of Müller (1992))

$$(nh)^{1/2}\left(\hat{\Delta}(\hat{\tau}) - \Delta\right) \to_d N\left(0, 2\sigma^2\int K_-^2(v)dv\right).$$

We can see that when $\mathbf{E}[Y|X = x] = g(x)$ is continuous in $x$, (i.e. $\Delta(x) = 0$ for all $x$), the estimator $\hat{\tau}$ will diverge. Therefore, the jump location estimator developed by Müller (1992) and Delgado and Hidalgo (2000) does not immediately apply when the presence of jump(s) in the regression function is unknown. In incomplete information games, the conditional choice probabilities *may* or *may not* have a jump in $x$, depending on the equilibrium behaviour. As a result, the econometrician cannot utilize the jump location estimator before knowing the presence of jump(s). This calls for a test for the presence of jump(s) in the conditional choice probabilities in the first place.

## 3.5 Testing for the Presence of Jump(s)

In nonparametric estimation, the regression function is usually assumed to be twice continuously differentiable. However, we have shown that even when the payoff functions and the latent distribution are all smooth, the equilibrium behaviour may give rise to jump(s) in the conditional choice probabilities. Hence the standard smoothness conditions imposed on regression functions are too restrictive for the conditional choices probabilities. There are generally two approaches to treat this problem, the first one is to conduct a testing procedure for the presence of jump(s) in the conditional choice probabilities. If the conditional probabilities are continuous in covariates, the standard smooth nonparametric estimators apply. Otherwise, the econometrician may estimate the locations of jumps (for example, use the method by Müller (1992) and Delgado and Hidalgo (2000)). An alternative approach is to estimate the conditional probability without the smoothness conditions (for example, use wavelet methods). In this section, we discuss the first approach, that is, to test for the presence of a jump in conditional choice probabilities. Since the presence of a jump is related to the equilibrium behaviour of the game, such a test not only helps the econometrician choose appropriate estimation approaches, but also reveals information about the equilibrium behaviour of the game. Suppose we are interested in testing

$$\begin{aligned} H_0' &: \ Q_k\left(1|1,x\right) \text{ is continuous in } x \text{ on } \mathrm{int}(\mathcal{X}) \text{ , for all } k=1,2. \\ H_1' &: \ Q_k\left(1|1,x\right) \text{ has jump(s) at some } x \in \mathrm{int}(\mathcal{X}), \text{ for some } k \in \{1,2\}. \end{aligned}$$

Since there are two players, the null hypothesis contains two sub-hypotheses. Taking this into account, if we focus on the conditional probability for each of the players, the critical value is to be adjusted (using Bonferroni correction, for example). In the following, we focus on testing the presence of a jump in the conditional choice probability for player 1.

The observed data is an i.i.d. sample $\{(Y_{1i}, Y_{2i}, X_i'\}_{i=1}^N$. Let $J = \dim(\mathcal{X})$. When $J = 1$, for a fixed $x$, let $\hat{g}_1^+(x)$ and $\hat{g}_1^-(x)$ be two kernel type estimators of $Q_1\left(1|1,x\right)$ using one-sided kernel functions[22], an estimator for the local jump at $x$ can be written as

$$\Delta_n(x) = \hat{g}_1^+(x) - \hat{g}_1^-(x),$$

where

$$\hat{g}^+(x) = \frac{\sum_{i=1}^N Y_{1i}Y_{2i}K_+\left(\frac{X_i-x}{h_n}\right)}{\sum_{i=1}^N Y_{2i}K_+\left(\frac{X_i-x}{h_n}\right)} \text{ and } \hat{g}^-(x) = \frac{\sum_{i=1}^N Y_{1i}Y_{2i}K_-\left(\frac{X_i-x}{h_n}\right)}{\sum_{i=1}^N Y_{2i}K_-\left(\frac{X_i-x}{h_n}\right)} \ ,$$

where $K_+(\cdot)$ and $K_-(\cdot)$ are one-sided kernel functions and $h_n$ is the bandwidth. The next step is to construct a test statistic by aggregating the local jumps over the support. The most straightforward way is to use supremum. To find the critical value [23], one can use the result from Hamrouni (1999)[24]. There are other ways to aggregate the local jumps. For instance, Bowman, Pope and Ismail (2006) developed a test based on the sum of squared pointwise jumps for the univariate and bivariate

[22]Properties of the one sided-kernel functions for $J = 1$ can be found in Definition 1.1 of Hamrouni (1999).

[23]Here one cannot directly apply the Continuous Mapping Theorem to deal with the supremum, because the underlying empirical process of kernel estimators does not have weak convergence.

[24]See Theorem 2.8 of Hamrouni (1999) for a complete statement of the limiting distribution.

regression functions. Alternatively, Müller and Stadtmüller (1999) showed that in univariate regression models, the sum of squared jump sizes can be represented as a coefficient of an asymptotic linear model, with the squared difference in $Y$ as the dependent variable and twice the standard deviation of the error term as the intercept. Hence they developed a test by checking whether the sum of the squared jump sizes equals zero. Gijbels and Goderniaux (2004) identified a discontinuity as a point with the largest derivative. When $J > 1$, the local jump at a given $x$ shall be checked along different directions, see Qiu (1997, 2002). A recent development in the literature related to this testing problem lies in Chernozhukov, Chetverikov and Kato (2012, 2013). They show that the supremum of an empirical process of kernel type estimators can be approximated by the supremum of a Gaussian multiplier bootstrap. This technique can be applied to our set-up. The details of constructing a test for the presence of jump(s) in the conditional choice probabilities in incomplete information games is left for future work.

## 3.6 Conclusions and Remarks

In this chapter, we show that in binary games with incomplete information, the conditional choice probabilities may have a jump with respect to the continuous covariates, even when the payoff functions and the distribution of the latent variables are all smooth. The source of such jumps lies in the equilibrium behaviour of the game. We further analyse the conditions under which the conditional choice probabilities exhibit a jump. The relationship between the equilibrium behaviour and the the presence of jump(s) in the conditional choice probabilities is robust to correlated private information and unobserved heterogeneity. As a result, continuity or discontinuity in the conditional choice probabilities reveals information about the equilibrium behaviour.

Testing for the presence of a jump in the conditional choice probabilities also matters if the econometrician wants to estimate that conditional choice probability, which is usually the starting point for parameter estimation and inference in incomplete information games. This chapter points out that when the econometrician is unknown about the presence of jump(s) in the conditional choice probability, the jump location estimator (by Müller (1992) and Delgado and Hidalgo (2000)) may not be consistent.

Lastly, we briefly discuss the construction of a test for the presence of a jump in the conditional choice probabilities. The test statistic can be constructed as the supremum of the local jumps, where the local jump can be computed as the difference between two kernel type estimators (using one-sided kernels). We leave the detailed construction of the test statistic and critical values for future work.

# Chapter 4

# Efficient Inference in Econometric Models When Identification Can Be Weak

## 4.1 Introduction

Weak instrumental variables (Weak-IV) have received a lot of attention in econometrics in particular following Staiger and Stock (1997), who developed an analytical framework for analysing the effect of weak instruments and constructing weak-identification-robust methods of inference, and Dufour (1997), who showed that usual bounded confidence sets cannot be valid in the case of weak identification. This paper considers testing the coefficient of an endogenous variable in instrumental variables regression models (linear IV models) when the instruments may be weak. A standard approach to the construction of Weak-IV robust tests is to use null-restricted residuals, which are obtained by imposing the value specified under a null hypothesis for the coefficients of endogenous regressors. The hypothesis can be tested by considering the sample covariance between null-restricted residuals and instrumental variables. The approach is robust to weak identification problems because, under a null hypothesis, the distribution of the covariance term does not depend asymptotically on the strength of the correlation between endogenous regressors and instruments. This idea is behind the Anderson-Rubin (AR) statistic, see Anderson and Rubin (1949) and Staiger and Stock (1997).

Tests based on the AR statistic are efficient if models are just identified. However, the power of AR-type tests is inferior to that of usual t- and Wald-tests when models are overidentified and instruments are strong. This is because, when models are overidentified, the AR approach tests more restrictions than the dimension of the parameter of interest. To address that issue, several papers suggested alternatives to the AR statistic. Kleibergen (2002, 2007) and Moreira (2001, 2003) proposed

Lagrange Multiplier (LM) and Conditional Likelihood Ratio (CLR) -type statistics. Kleibergen's LM (KLM) tests can be used with usual $\chi^2$ critical values. CLR tests requires simulations to generate critical values, however, they have been demonstrated to have better power properties in Monte Carlo simulations than KLM or AR tests. Andrews, Moreira and Stock (2006) showed that, for normal linear instrumental variables regression models with homoskedastic errors, KLM and CLR tests are efficient among Weak-IV robust tests when instrumental variables are strong (Strong-IV). When instrumental variables are weak, there is no uniformly most powerful (UMP) test except for the just identified case, hence the efficiency depends on optimizing criteria considered by the econometrician. Andrews, Moreira and Stock (2006) considered a weighted average power (WAP) at two values of parameters, and derived the optimal average power test. They also numerically demonstrated that in the Weak-IV scenario, CLR test dominates AR and KLM tests, and attains the power envelope given by the optimal average power test. As a result, they recommended CLR test for empirical researchers when the instruments may be weak and model is over-identified. Cattaneo, Crump and Jansson (2012) extended the results of Andrews, Moreira and Stock (2006) to non-normal errors by using an asymptotic framework of Gaussian experiments. Chernozhukov, Hansen and Jansson (2009) showed that all members of the weighted average power likelihood ratio tests are admissible, including the AR test.

This chapter is concerned with a different optimizing criterion, that is, the power for alternative hypotheses that are determined by arbitrarily large deviations from the null hypothesis. In the Weak-IV scenario, the power of any robust test may be far below 1 even for arbitrarily large deviations. The power of a test for such alternatives is also related to the length of confidence intervals constructed using test inversion.

In this chapter, we focus on an asymptotic experiment following Cattaneo, Crump and Jansson (2012) and Choi and Schick (1996). This asymptotic experiment framework substantially simplifies the analysis by reducing a complex inference problem to that based on a normally distributed vector. It allows one to derive an efficiency bound in presence of nuisance parameters. In this chapter, we first derive the optimal test for alternatives that are arbitrarily far away from the null, among all rotational invariant and asymptotically similar tests. To do this, we follow the method of Andrews, Moreira and Stock (2006) and Mills, Moreira and Vilela (2013), with a focus on alternatives determined by arbitrarily large deviations from the null. Then we use the notion of efficiency in Choi and Schick (1996) to obtain a power envelope under Weak-IV, in the worst scenario with respect to a perturbation to the nuisance parameter. After that, we compare the power of popular Weak-IV robust tests (AR, KLM and CLR tests) when alternatives are determined by arbitrarily large deviations from the null. In particular, we find that the relative performance of the AR test versus the CLR test depends on the degree of endogeneity in the model. For a relatively low degree of endogeniety, the AR test outperforms the CLR test, while for a relatively large degree of endogeniety, the order is reversed. This result suggests that the CLR test is not dominating the AR test under a different (but reasonable) optimizing criterion from the WAP considered by Andrews, Moreira and Stock (2006). In addition, we propose a new Weak-IV robust test, the Conditional Lagrange Multiplier (CLM) test, which is asymptotically efficient in the Strong-IV case, robust to weak instruments, and exhibits the same power as the AR test for arbitrarily large deviations from the null. Lastly, we extend the investigation to heteroskedastic models. In particular, we find that the generalized likelihood ratio statistic in heteroskedastic models reduces to the AR statistic in the Weak-IV scenario and when alternatives are determined by arbitrarily large deviations from the null.

### 4.1.1 Organization of Chapter 4

The plan of the chapter is as follows: Section 4.2 sets up the asymptotic experiment framework, Section 4.3 describes the optimal rotational invariant and asymptotically similar test when alternatives are determined by arbitrarily deviations from the null; Section 4.4 characterizes the power envelope in the worst scenario with respect to a perturbation to the nuisance parameter. Section 4.5 compares the power properties of the AR, CLR and KLM tests under Weak-IV and when alternatives are determined by arbitrarily large deviations from the null. This section also proposes a new Weak-IV robust test, the CLM test. Section 4.6 extends the framework to heteroskedastic models. The last section concludes.

## 4.2 An Asymptotic Experiment for Linear IV Models

Consider linear IV models with a single endogenous regressor. The structural equation is

$$y_1 = y_2\gamma + Z_2\beta + u, \tag{4.1}$$

and the first stage regression is

$$y_2 = Z_1\pi_{1,n} + Z_2\pi_2 + v, \tag{4.2}$$

where $y_1, y_2 \in \mathbb{R}^n$, $Z_2 \in \mathbb{R}^{n\times l_1}$, and $Z_2 \in \mathbb{R}^{n\times l_2}$ are observed variables; $u, v \in \mathbb{R}^n$ are unobserved error terms; coefficients $\gamma \in \mathbb{R}$, $\beta, \pi_2 \in \mathbb{R}^{l_2}$, and $\pi_{1,n} \in \mathbb{R}^{l_1}$ are unknown parameters, among which the coefficient $\gamma$ is the structural parameter of interest. Assumption 1 characterizes the Weak-IV scenario.

**Assumption 1. (Weak-IV)** $\pi_{1,n} = n^{-1/2}C$, where $C \in \mathbb{R}^{l_1}$ is fixed.

We assume that the data is an i.i.d. sample, the instrumental variables are uncorrelated with the unobserved error terms, the errors in the model are homoskedastic, and the instrumental variables have finite second order moments.

**Assumption 2.** (a) $\{(y_{1i}, y_{2i}, Z_{1i}, Z_{2i}), i = 1, ..., n\}$ are i.i.d.

(b) $\mathrm{E}\begin{bmatrix} Z_{1i} \\ Z_{2i} \end{bmatrix}\begin{bmatrix} u_i & v_i \end{bmatrix} = 0.$

(c) $\mathrm{E}\left[\begin{matrix} u_i^2 & u_iv_i \\ u_iv_i & v_i^2 \end{matrix}\ |Z_{i1}, Z_{i2}\right] = \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix}$, is a finite and positive definite matrix.

(d) $\mathrm{E}\begin{bmatrix} Z_{1i}Z_{1i}' & Z_{1i}Z_{2i}' \\ Z_{2i}Z_{1i}' & Z_{2i}Z_{2i}' \end{bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}' & Q_{22} \end{bmatrix} = Q$, is a finite and positive definite matrix.

By Assumption 2(c) and the fact that $\frac{Z_1'M_2Z_1}{n} \to_p Q_{1\cdot 2}$, we have

$$\frac{1}{\sqrt{n}}\begin{bmatrix} Z_1'M_2u \\ Z_1'M_2v \end{bmatrix} \to_d N\left(0, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix} \otimes Q_{1\cdot 2}\right), \tag{4.3}$$

where $Q_{1\cdot 2} = Q_{11} - Q_{12}Q_{22}^{-1}Q_{12}'$. Note that (4.3) is a characterization of homoskedasticity.

The null hypothesis is $H_0 : \gamma = \gamma_0$. Let $\Delta = \gamma - \gamma_0$. The following two statistics $S_n$ and $T_n$ (and their normalized versions $S_n^*$ and $T_n^*$) will be used repeatedly in this chapter. We construct the statistic $S_n \in \mathbb{R}^{l_1}$ as the sample covariance between the null restricted residuals and the instrumental variables.

$$S_n \;=\; Z_1'M_2(y_1 - y_2\gamma_0)/n$$

$$= \frac{Z_1' M_2 Z_1}{n} \Delta \frac{C}{\sqrt{n}} + \frac{Z_1' M_2 (u + \Delta v)}{n}, \tag{4.4}$$

where $M_2 = I_n - Z_2 (Z_2' Z_2)^{-1} Z_2'$ is an orthogonal projection matrix. Note that the second equality is implied by Assumption 1. Let $\hat{\pi}_{1,n}$ be the OLS estimator of $\pi_{1,n}$ in the first stage regression, that is

$$\hat{\pi}_{1,n} = (Z_1' M_2 Z_1)^{-1} Z_1' M_2 y_2 = \left( \frac{Z_1' M_2 Z_1}{n} \right)^{-1} \frac{Z_1' M_2 v}{n} + \frac{C}{\sqrt{n}}, \tag{4.5}$$

where the second equality comes from Assumption 1. Define

$$\sigma^2(\Delta) = \sigma_u^2 + \Delta^2 \sigma_v^2 + 2\Delta \sigma_{uv}.$$

Expressions in (4.4) and (4.5) give rise to the asymptotic distribution of $\sqrt{n}[S_n', \hat{\pi}_{1,n}']'$ in the Weak-IV scenario:

$$N \left( \begin{bmatrix} \Delta Q_{1\cdot 2} C \\ C \end{bmatrix}, \begin{bmatrix} \sigma^2(\Delta) Q_{1\cdot 2} & \left(\sigma_{uv} + \Delta \sigma_v^2\right) I_{l_1} \\ \left(\sigma_{uv} + \Delta \sigma_v^2\right) I_{l_1} & \sigma_v^2 Q_{1\cdot 2}^{-1} \end{bmatrix} \right).$$

Next we decompose $\hat{\pi}_{1,n}$ into two parts, one is the population projection of $\hat{\pi}_{1,n}$ onto the space of $S_n$ and the other is orthogonal in population to $S_n$. We take the second part and construct the statistic $T_n \in \mathbb{R}^{l_1}$ to be asymptotically uncorrelated with $S_n$:

$$T_n = \hat{\pi}_{1,n} - \tfrac{\sigma_{uv} + \Delta \sigma_v^2}{\sigma^2(\Delta)} Q_{1\cdot 2}^{-1} S_n.$$

The asymptotic distribution of$\sqrt{n}$ $[S_n', T_n']'$ is

$$N \left( \begin{bmatrix} \Delta Q_{1\cdot 2} C \\ \frac{\sigma_u^2 + \Delta \sigma_{uv}}{\sigma^2(\Delta)} C \end{bmatrix}, \begin{bmatrix} \left(\sigma_u^2 + 2\Delta\sigma_{uv} + \Delta^2 \sigma_{vv}\right) Q_{1\cdot 2} & 0_{l_1} \\ 0_{l_1} & \frac{\sigma_u^2 \sigma_v^2 - \sigma_{uv}^2}{\sigma^2(\Delta)} Q_{1\cdot 2}^{-1} \end{bmatrix} \right).$$

We normalize $S_n$ and $T_n$ by letting

$$S_n^* = Q_{1\cdot 2}^{-1/2} \sqrt{n} S_n / \sigma(\Delta),$$

and

$$T_n^* = \sigma(\Delta) Q_{1\cdot 2}^{1/2} \sqrt{n} T_n / (\sigma_u^2 \sigma_v^2 - \sigma_{uv}^2)^{1/2}.$$

Since the covariance matrix of the reduced-form errors $[u_i + \Delta v_i, v_i]'$ can be consistently estimated, in this chapter we assume that $Q_{1\cdot 2}$, $\sigma(\Delta)$ and $\sigma_u^2 \sigma_v^2 - \sigma_{uv}^2$ are known. Note that $Q_{1\cdot 2}$ can be consistently estimated by

$\frac{Z_1' M_2 Z_1}{n}$. Let $\Omega$ denote the 2 by 2 covariance matrix of the reduced form [25] errors $[u_i + \Delta v_i, v_i]'$. Then $\sigma(\Delta)$ is the upper left element of $\Omega$, $\sigma_{uv} + \Delta\sigma_v^2$ is the upper right element of $\Omega$, and $\sigma_u^2\sigma_v^2 - \sigma_{uv}^2$ is the determinant of $\Omega$.

Furthermore, we obtain the following asymptotic experiment

$$\begin{bmatrix} S_n^* \\ T_n^* \end{bmatrix} \sim_a N\left(\begin{bmatrix} \frac{\Delta}{\sigma(\Delta)} Q_{1\cdot 2}^{1/2} C \\ \frac{(\sigma_u^2 + \Delta\sigma_{uv})}{\sigma(\Delta)(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2)^{1/2}} Q_{1\cdot 2}^{1/2} C \end{bmatrix}, I_{2l_1}\right). \tag{4.6}$$

Define $\mathcal{N}$ to be a $2l_1 \times 1$ normal random vector with a zero mean and the identity covariance matrix, and let

$$\begin{bmatrix} S^* \\ T^* \end{bmatrix} = \mathcal{N} + \begin{bmatrix} \frac{\Delta}{\sigma(\Delta)} Q_{1\cdot 2}^{1/2} C \\ \frac{(\sigma_u^2 + \Delta\sigma_{uv})}{\sigma(\Delta)(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2)^{1/2}} Q_{1\cdot 2}^{1/2} C \end{bmatrix}.$$

Clearly, $\begin{bmatrix} S^* \\ T^* \end{bmatrix}$ is distributed as the right hand side of (4.6), and we have

$$\begin{bmatrix} S_n^* \\ T_n^* \end{bmatrix} \to_d \begin{bmatrix} S^* \\ T^* \end{bmatrix}.$$

We are particularly interested in alternatives that are determined by arbitrarily large deviations from the null, that is, $\Delta \to \infty$. In the following investigation, we first analyse the case $\Delta \to +\infty$, and the case $\Delta \to -\infty$ can be treated analogously. Consider the asymptotic experiments under the null and alternatives, under $H_0: \ \Delta = 0$,

$$\begin{bmatrix} S^* \\ T^* \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ \frac{1}{\sqrt{1-\rho^2}}\lambda \end{bmatrix}, I_{2l_1}\right), \tag{4.7}$$

where $\lambda = \frac{1}{\sigma_v} Q_{1\cdot 2}^{1/2} C$ and $\rho = \frac{\sigma_{uv}}{\sigma_u\sigma_v}$.

Under $H_1: \Delta \to +\infty$,

$$\begin{bmatrix} S^* \\ T^* \end{bmatrix} \sim N\left(\begin{bmatrix} \lambda \\ \frac{\rho}{\sqrt{1-\rho^2}}\lambda \end{bmatrix}, I_{2l_1}\right), \tag{4.8}$$

[25] Here the reduced form equations are

$$\begin{bmatrix} y_1 - \gamma_0 y_2 \\ y_2 \end{bmatrix} = Z_1 \begin{bmatrix} \Delta\pi_{1,n} \\ \pi_{1,n} \end{bmatrix} + Z_2 \begin{bmatrix} \Delta\pi_2 + \beta \\ \pi_2 \end{bmatrix} + \begin{bmatrix} u + \Delta v \\ v \end{bmatrix}.$$

Similarly, under $H_1 : \ \Delta \to -\infty$,

$$\left[ \begin{array}{c} S^* \\ T^* \end{array} \right] \sim N\left( - \left[ \begin{array}{c} \lambda \\ \frac{\rho}{\sqrt{1-\rho^2}}\lambda \end{array} \right], I_{2l_1} \right). \tag{4.9}$$

Asymptotic experiments (4.7), (4.8) and (4.9) are the building blocks of this chapter. We can see that the means of those limiting Gaussian distributions are determined by three parameters: $l_1$, $\lambda$ and $\rho$. The parameter $l_1$ is the number of instruments; the norm of $\lambda$ determines the strength of the instruments; and the parameter $\rho$ measures the degree of endogeneity of this linear IV model.

## 4.3 The Optimal Rotational Invariant and Asymptotically Similar Test

In this section, we derive the asymptotically optimal test against $\Delta \to \infty$ among all rotational invariant and asymptotically similar tests. Here rotational invariance means that a test is not affected by any orthonormal transformation of the instruments. Asymptotic similarity means that the asymptotic null rejection rate of a test is not affected by $\pi_{1,n}$. Similar tests are robust to weak instruments, as the norm of $\pi_{1,n}$ determines the strength of the instruments. Define the test statistics

$$\begin{aligned} Q_n &= \left[ \begin{array}{cc} Q_{sn} & Q_{stn} \\ Q'_{stn} & Q_{tn} \end{array} \right] = [S_n^*, T_n^*]'[S_n^*, T_n^*], \\ Q &= \left[ \begin{array}{cc} Q_s & Q_{st} \\ Q'_{st} & Q_t \end{array} \right] = [S^*, T^*]'[S^*, T^*]. \end{aligned}$$

In this following, we consider test statistics that are functions of $Q_n$. This is because Andrews, Moreira and Stock (2006) showed that every rotational invariant test can be written as a function of $Q_n$. In addition, they showed that an invariant test is asymptotically similar with significance level $\alpha$ if and only if the asymptotic null rejection rate of such a test equals $\alpha$, conditional on the value of $Q_{tn}$. Therefore, we can restrict our attention to tests as functions of $Q_n$. Popular Weak-IV robust tests, such as the AR, KLM and CLR tests, can be written as functions of $Q_n$.

From the asymptotic experiment (4.6), the $l_1 \times 2$ random matrix $[S^*, T^*]$ is a multivariate normal with the mean matrix given by

$$M = \left[ \begin{array}{cc} \frac{\Delta}{\sigma(\Delta)}, & \frac{(\sigma_u^2 + \Delta\sigma_{uv})}{\sigma(\Delta)(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2)^{1/2}} \end{array} \right] Q_{1\cdot 2}^{1/2} C,$$

and the identity covariance matrix. The $2 \times 2$ random matrix $Q$ has a non-central Wishart distribution with the mean matrix of rank 1 and the identity covariance matrix. Therefore we can calculate the density of $[Q_s, Q_{st}]$ conditional on $Q_t$, and then use the Neyman-Pearson Lemma to construct the optimal test against $H_1 : \Delta \to +\infty$ or $\Delta \to -\infty$. This method has been used by Andrews, Moreira and Stock (2006) and Mills, Moreira and Vilela (2013). Here we particularly focus on alternatives corresponding to arbitrarily large deviations from the null. The following proposition describes the optimal rotational invariant and asymptotically similar test in the Weak-IV scenario and for alternatives corresponding to arbitrarily large deviations from the null.

**Proposition 1.** *In the linear IV model (4.1) and (4.2), suppose that Assumption 1 and 2 hold. Consider a testing problem*

$$H_0 : \ \Delta = 0, \ H_1 : \Delta \to \infty.$$

*Then the test that rejects $H_0$ when*

$$POIS_\infty(Q_{sn}, Q_{stn}) = Q_{sn} + \frac{2\rho}{\sqrt{1-\rho^2}} Q_{stn} > \kappa_\infty(Q_{tn})$$

*maximizes the asymptotic power over all tests as functions of $Q_n$ and with asymptotic size $\alpha$, where $\kappa_\infty(Q_{tn})$ is the $(1-\alpha)$th quantile of the distribution of $POIS_\infty(Q_{sn}, Q_{stn})$ conditional on $Q_{tn}$ and under $H_0$.*

*Proof.* We first characterize the optimal invariant and similar test based on the asymptotic statistic $Q$. If the optimal test statistic and the critical value are continuous in $Q$, then by the Continuous Mapping Theorem, replacing $Q$ with $Q_n$ yields a test that attains the asymptotic power envelope among all invariant and asymptotically similar tests.

The density of $(Q_s, Q_{st}, Q_t)$ is

$$\begin{aligned} f_Q(q_s, q_{st}, q_t) &= K_1 \exp\left(-\mathrm{tr}\left(M'M\right)/2\right) \det(q)^{(l_1-3)/2} \exp\left(-\mathrm{tr}\left(q\right)/2\right) \\ &\quad \times \mathrm{tr}\left(M'Mq\right)^{-(l_1-2)/4} I_{(l_1-2)/2} \sqrt{\mathrm{tr}\left(M'Mq\right)}, \end{aligned}$$

where

$$I_v(x) = \left(\frac{x}{2}\right)^v \sum_{j=0}^{\infty} \frac{(x^2/4)^j}{j!\Gamma(v+j+1)},$$

and $K_1$ is a constant only depending on $l_1$. Calculate the mean matrices under $H_0$ and $H_1$, respectively.

$$\text{Under } H_0 : \Delta = 0, \quad M \ = [0, \frac{1}{\sqrt{1-\rho^2}}\lambda]',$$

$$\text{Under } H_1 : \Delta \to +\infty, \quad M \quad = [\lambda, \frac{\rho}{\sqrt{1-\rho^2}}\lambda].$$

Therefore, under $H_0$, the density of $(Q_s, Q_{st}, Q_t)$ is

$$\begin{aligned} f_Q^0(q_s, q_{st}, q_t) \quad &= \quad K_1 \exp\left(-\frac{\|\lambda\|^2}{2(1-\rho^2)}\right) \det(q)^{\frac{l_1-3}{2}} \exp\left(-\frac{q_s+q_t}{2}\right) \\ &\times \left(\|\lambda\|^2 \frac{q_t}{(1-\rho^2)}\right)^{-\frac{l_1-2}{4}} I_{\frac{l_1-2}{2}}\left(\sqrt{\|\lambda\|^2 \frac{q_t}{(1-\rho^2)}}\right). \end{aligned}$$

Under $H_1 : \Delta \to +\infty$, the density of $(Q_s, Q_{st}, Q_t)$ becomes

$$\begin{aligned} f_Q^1(q_s, q_{st}, q_t) \quad &= \quad K_1 \exp\left(-\frac{\|\lambda\|^2}{2(1-\rho^2)}\right) \det(q)^{\frac{l_1-3}{2}} \exp\left(-\frac{q_s+q_t}{2}\right) \\ &\times \left(\|\lambda\|^2 \zeta(q)\right)^{-\frac{l_1-2}{4}} I_{\frac{l_1-2}{2}}\left(\sqrt{\|\lambda\|^2 \zeta(q)}\right), \end{aligned}$$

where

$$\zeta(q) = q_s + \frac{2\rho}{\sqrt{1-\rho^2}} q_{st} + \frac{\rho^2}{1-\rho^2} q_t.$$

The random variable $Q_t$ has a non-central chi-square distribution, and the densities of $Q_t$ under $H_0$ and $H_1$ can be written as

$$\begin{aligned} f_{Q_t}^0(q_t) \quad &= \quad \frac{1}{2} \exp\left(-\frac{\|\lambda\|^2}{2(1-\rho^2)}\right) q_t^{\frac{l_1-2}{2}} \exp\left(-\frac{q_t}{2}\right) \\ &\times \left(\|\lambda\|^2 \frac{q_t}{(1-\rho^2)}\right)^{-\frac{l_1-2}{4}} I_{\frac{l_1-2}{2}}\left(\sqrt{\|\lambda\|^2 \frac{q_t}{(1-\rho^2)}}\right), \end{aligned}$$

and

$$\begin{aligned} f_{Q_t}^1(q_t) \quad &= \quad \frac{1}{2} \exp\left(-\frac{\rho^2\|\lambda\|^2}{2(1-\rho^2)}\right) q_t^{\frac{l_1-2}{2}} \exp\left(-\frac{q_t}{2}\right) \\ &\times \left(\|\lambda\|^2 \frac{\rho^2 q_t}{(1-\rho^2)}\right)^{-\frac{l_1-2}{4}} I_{\frac{l_1-2}{2}}\left(\sqrt{\|\lambda\|^2 \frac{\rho^2 q_t}{(1-\rho^2)}}\right). \end{aligned}$$

Therefore, we can compute the densities of $(Q_s, Q_{st})$ conditional on $Q_t$, under $H_0$ and $H_1$ respectively. As a result, the likelihood ratio can be written as

$$\begin{aligned}
LR(q) &= \frac{f_Q^1(q_s, q_{st}, q_t)/f_{Q_t}^1(q_t)}{f_Q^0(q_s, q_{st}, q_t)/f_{Q_t}^0(q_t)} \\
&= \frac{\exp\left(-\frac{\|\lambda\|^2}{2}\right)\left(\|\lambda\|^2\zeta(q)\right)^{-\frac{l_1-2}{4}} I_{\frac{l_1-2}{2}}\left(\sqrt{\|\lambda\|^2\zeta(q)}\right)}{\left(\|\lambda\|^2\frac{q_t}{(1-\rho^2)}\right)^{-\frac{l_1-2}{4}} I_{\frac{l_1-2}{2}}\left(\sqrt{\|\lambda\|^2\frac{\rho^2 q_t}{(1-\rho^2)}}\right)} \\
&= \frac{2^{-\frac{l_1-2}{2}}\exp\left(-\frac{\|\lambda\|^2}{2}\right)\sum_{j=0}^{\infty}\frac{(\|\lambda\|^2\zeta(q)/4)^j}{j!\Gamma(l_1/2+j)}}{\left(\|\lambda\|^2\frac{q_t}{(1-\rho^2)}\right)^{-\frac{l_1-2}{4}} I_{\frac{l_1-2}{2}}\left(\sqrt{\|\lambda\|^2\frac{\rho^2 q_t}{(1-\rho^2)}}\right)}. \qquad (4.10)
\end{aligned}$$

Note that the denominator of (4.10) is a function of $q_t$ and the numerator is an increasing function of $\zeta(q)$. Therefore, by the Neyman-Pearson Lemma, the optimal test for $H_1 : \Delta \to +\infty$ is to reject $H_0$ when

$$POIS_\infty(Q_s, Q_{st}) = Q_s + \frac{2\rho}{\sqrt{1-\rho^2}}Q_{st} > \kappa_\infty(Q_t),$$

where the critical value $\kappa_\infty(Q_t)$ is the $(1-\alpha)$th quantile of the distribution of $POIS_\infty(Q_s, Q_{st})$ conditional on $Q_t$ and under $H_0$.

Since both $POIS_\infty(\cdot)$ and $\kappa_\infty(\cdot)$ are continuous functions of $Q$, by the Continuous Mapping Theorem, the test that rejects $H_0$ when

$$POIS_\infty(Q_{sn}, Q_{stn}) = Q_{sn} + \frac{2\rho}{\sqrt{1-\rho^2}}Q_{stn} > \kappa_\infty(Q_{tn}),$$

maximizes the asymptotic power over all tests as functions of $Q_n$ and with asymptotic size $\alpha$, where $\kappa_\infty(Q_{tn})$ is the $(1-\alpha)$th quantile of the distribution of $POIS_\infty(Q_{sn}, Q_{stn})$, conditional on $Q_{tn}$ and under $H_0$.

Exactly the same result can be obtained for $H_1 : \Delta \to -\infty$. □

The optimal test given by Proposition 1 does not depend on the nuisance parameter $C$, which cannot be consistently estimated in the Weak-IV scenario. However, the optimal test is still infeasible because $\rho$ is unknown and cannot be consistently estimated in linear IV models. The parameter $\rho$ measures the degree of endogeneity in the model.

Mills, Moreira and Vilela (2013) proposed a feasible test for arbitrarily large $\Delta$. Their test is to reject $H_0$ when

$$Q_{sn} + 2(\det(\Omega))^{-1/2}(\gamma_0\omega_{22} - \omega_{12})Q_{stn} > \kappa_\infty(Q_{tn}),$$

where the $\kappa_\infty(Q_{tn})$ is the $(1-\alpha)$th quantile of null distribution of the left hand side conditional on $Q_{tn}$, and $\Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{bmatrix}$ is the covariance matrix of the reduced-form errors $\begin{bmatrix} u+\gamma v \\ v \end{bmatrix}$. However, when $\gamma \to \infty$, $\gamma_0\omega_{22} - \omega_{12} = -\Delta\sigma_v^2 - \sigma_{uv}$, which goes to infinity. Therefore, the behaviour of the test statistic is determined by the second term $Q_{stn}$. A test based on $Q_{stn}$ will exhibit some undesirable power properties similar to Kleibergen's K (KLM) test.[26] The power performance of KLM will be illustrated in Section 4.5.

## 4.4 The Power Envelope Under an Unknown Nuisance Parameter

In this section, we derive a power envelope in the worst scenario with respect to a perturbation to the nuisance parameter $C$, in the Weak-IV scenario. Consider a perturbation to the nuisance parameter $C$, i.e., $C = C_0 + \tau$. We will be focusing on the asymptotic power envelope for a joint test of

$$H_0 : \Delta = 0 \text{ and } \tau = 0,$$

against

$$H_1 : \Delta \to \infty \text{ and } \tau \neq 0.$$

The following proposition characterizes the power envelope in that case.

**Proposition 2.** *In the linear IV model (4.1),(4.2), suppose that Assumption 1 and 2 hold. Consider the testing problem*

$$H_0 : \Delta = 0 \textit{ and } \tau = 0,$$

*against*

$$H_1 : \Delta \to \infty \textit{ and } \tau \neq 0.$$

*Then the asymptotic power envelope in the worst scenario with respect to $\tau$ for tests with an asymptotic size $\alpha$ is*

$$Pr\left(G^* > \chi^2_{1,1-\alpha}\right), \textit{ where } G^* \sim \chi^2_1\left(\frac{C_0' Q_{1\cdot 2} C_0}{\sigma_v^2}\right).$$

[26] Mills, Moreira and Vilela (2013) first computed the optimal test for a fixed $\gamma$, and then sent $\gamma$ to infinity. However, $\omega_{12}$ is a function of $\gamma$ and will also go to $\infty$. This is not taken into account in their test statistic.

*That is, the optimal test statistic in the worst scenario with respect to* $\tau$ *has a non-central chi-square limiting distribution with degree of freedom* 1 *and the non-centrality parameter equal to* $C_0' Q_{1\cdot 2} C_0 / \sigma_v^2$. *The critical value* $\chi^2_{1,1-\alpha}$ *is the* $(1-\alpha)$*th quantile of a centred chi-square distribution.*

*Proof.* For a fixed $(\Delta, \tau)$, the asymptotic experiment in (4.6) becomes

$$\begin{bmatrix} S^* \\ T^* \end{bmatrix} \sim N\left( \begin{bmatrix} \frac{\Delta}{\sigma(\Delta)} Q_{1\cdot 2}^{1/2} (C_0 + \tau) \\ \frac{(\sigma_u^2 + \Delta\sigma_{uv})}{\sigma(\Delta)(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2)^{1/2}} Q_{1\cdot 2}^{1/2} (C_0 + \tau) \end{bmatrix}, I_{2l_1} \right).$$

The likelihood ratio statistic for $H_1 : \Delta \neq 0$, $C = C_0 + \tau$ is

$$\begin{aligned} LR \;=\; & -\left(S^* - \tfrac{\Delta}{\sigma(\Delta)} Q_{1\cdot 2}^{1/2} (C_0 + \tau)\right)' \left(S^* - \tfrac{\Delta}{\sigma(\Delta)} Q_{1\cdot 2}^{1/2} (C_0 + \tau)\right) + S^{*\prime} S^* \\ & - \left(T^* - \frac{(\sigma_u^2 + \Delta\sigma_{uv}) Q_{1\cdot 2}^{1/2} (C_0 + \tau)}{\sigma(\Delta)(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2)^{1/2}}\right)' \left(T^* - \frac{(\sigma_u^2 + \Delta\sigma_{uv}) Q_{1\cdot 2}^{1/2} (C_0 + \tau)}{\sigma(\Delta)(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2)^{1/2}}\right) \\ & + \left(T^* - \frac{\sigma_u Q_{1\cdot 2}^{1/2} C_0}{(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2)^{1/2}}\right)' \left(T^* - \frac{\sigma_u Q_{1\cdot 2}^{1/2} C_0}{(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2)^{1/2}}\right). \end{aligned} \tag{4.11}$$

When $\Delta \to +\infty$, the expression in (4.11) becomes

$$LR_{+\infty} = 2W_{+\infty} + K,$$

where $K$ is a constant and

$$\begin{aligned} W_{+\infty} \;=\; & \frac{S^{*\prime} Q_{1\cdot 2}^{1/2} (C_0 + \tau)}{\sigma_v} \\ & + \left(T^* - \frac{\sigma_u Q_{1\cdot 2}^{1/2} C_0}{(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2)^{1/2}}\right)' \left(\frac{Q_{1\cdot 2}^{1/2} (\sigma_{uv}(C_0 + \tau) - \sigma_u\sigma_v C_0)}{\sigma_v(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2)^{1/2}}\right). \end{aligned}$$

We compute

$$\begin{aligned} \mathrm{E}\left[W_{+\infty}\right] \;=\; & \frac{(C_0 + \tau)' Q_{1\cdot 2} (C_0 + \tau)}{\sigma_v^2} \\ & + \frac{((C_0 + \tau)\sigma_{uv} - \sigma_u\sigma_v C_0)' \, Q_{1\cdot 2} \, ((C_0 + \tau)\sigma_{uv} - \sigma_u\sigma_v C_0)}{\sigma_v^2(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2)}, \end{aligned} \tag{4.12}$$

and $\mathrm{Var}\left[W_{+\infty}\right] = \mathrm{E}\left[W_{+\infty}\right]$. Thus we have

$$G_{+\infty} = W_{+\infty}^2 / \mathrm{Var}\left[W_{+\infty}\right] \sim \chi_1^2(\mathrm{E}\left[W_{+\infty}\right]).$$

The asymptotically optimal level $\alpha$ test against $H_1 : \Delta \to +\infty$ is to reject $H_0$ when

$$G_{n,+\infty} = W_{n,+\infty}^2 / \mathrm{Var}\left[W_{+\infty}\right] > \chi^2_{1,1-\alpha},$$

where

$$\begin{aligned} W_{n,+\infty} &= \frac{S_n^{*\prime} Q_{1\cdot 2}^{1/2}(C_0+\tau)}{\sigma_v} \\ &\quad + \left(T_n^* - \frac{\sigma_u Q_{1\cdot 2}^{1/2} C_0}{(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2)^{1/2}}\right)' \left(\frac{Q_{1\cdot 2}^{1/2}\left(\sigma_{uv}(C_0+\tau) - \sigma_u\sigma_v C_0\right)}{\sigma_v(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2)^{1/2}}\right). \end{aligned} \tag{4.13}$$

The asymptotic distribution of the statistic $G_{n,+\infty}$ is a non-central chi-square with degree of freedom 1 and the non-centrality parameter equal to $\mathrm{E}\left[W_{+\infty}\right]$ in (4.12). To consider the worst case, we minimize the non-centrality parameter in (4.12) with respect to $\tau$. The resulting power minimizing direction in terms of the disturbance to $C_0$ is

$$\tau = \frac{\sigma_{uv} - \sigma_u\sigma_v}{\sigma_u\sigma_v} C_0.$$

Plugging the worst $\tau$ back into (4.12) yields the non-centrality parameter associated with the optimal test in the worst scenario. The resulting non-centrality parameter is $\frac{C_0' Q_{1\cdot 2} C_0}{\sigma_v^2}$. Therefore, the power envelope for $\Delta \to +\infty$ in the worst scenario with respect to $\tau$ is

$$\Pr\left(G^* > \chi^2_{1,1-\alpha}\right), \text{ where } G^* \sim \chi^2_1\left(\frac{C_0' Q_{1\cdot 2} C_0}{\sigma_v^2}\right).$$

When $\Delta \to -\infty$, the likelihood ratio statistic can be reduced to

$$\begin{aligned} W_{-\infty} &= -\frac{S^{*\prime} Q_{1\cdot 2}^{1/2}(C_0+\tau)}{\sigma_v} \\ &\quad - \left(T^* - \frac{\sigma_u Q_{1\cdot 2}^{1/2} C_0}{(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2)^{1/2}}\right)' \left(\frac{Q_{1\cdot 2}^{1/2}\left(\sigma_{uv}(C_0+\tau) + \sigma_u\sigma_v C_0\right)}{\sigma_v(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2)^{1/2}}\right). \end{aligned}$$

This yields

$$\begin{aligned} \mathrm{E}\left[W_{-\infty}\right] &= \frac{(C_0+\tau)' Q_{1\cdot 2}(C_0+\tau)}{\sigma_v^2} \\ &\quad + \frac{\left(\sigma_{uv}(C_0+\tau) + \sigma_u\sigma_v C_0\right)' Q_{1\cdot 2}\left(\sigma_{uv}(C_0+\tau) + \sigma_u\sigma_v C_0\right)}{\sigma_v^2(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2)}, \end{aligned}$$

and $\mathrm{Var}\left[W_{-\infty}\right]=\mathrm{E}\left[W_{-\infty}\right]$. Consequently, the power minimizing direction is $\tau=-\frac{\sigma_{uv}+\sigma_u\sigma_v}{\sigma_u\sigma_v}C_0$ and the resulting non-centrality parameter associated with the optimal test for $\Delta\to-\infty$ remains $\frac{C_0'Q_{1\cdot 2}C_0}{\sigma_v^2}$. □

By plugging the minimizer $\tau=\frac{\sigma_{uv}-\sigma_u\sigma_v}{\sigma_u\sigma_v}C_0$ into $W_{n,+\infty}$ given by (4.13), we obtain a statistic

$$W_{n,+\infty}^{*}=\frac{\sigma_{uv}}{\sigma_u\sigma_v^2}S_n^{*\prime}Q_{1\cdot 2}^{1/2}C_0-\frac{(\sigma_u^2\sigma_v^2-\sigma_{uv}^2)^{1/2}}{\sigma_u\sigma_v^2}T_n^{*\prime}Q_{1\cdot 2}^{1/2}C_0+\frac{C_0Q_{1\cdot 2}C_0}{\sigma_v^2}. \tag{4.14}$$

The asymptotically optimal test for $\Delta\to+\infty$, in the worst scenario with respect to $\tau$ is to reject $H_0$ when

$$G_{n,+\infty}^{*}=\frac{\sigma_v^2(W_{n,+\infty}^{*})^2}{C_0Q_{1\cdot 2}C_0}>\chi^2_{1,1-\alpha}. \tag{4.15}$$

For the alternative $H_1:\Delta\to-\infty$, we obtain exactly the same asymptotically optimal test in the worst scenario with respect to $\tau$. Note that the asymptotically optimal test is $C_0$-specific. However, $C_0$ cannot be consistently estimated. Hence the above asymptotically optimal test is infeasible.

## 4.5 Power Comparisons of Robust Tests

In this section, we evaluate the power of several popular Weak-IV robust tests, when alternatives are determined by arbitrarily large deviations from the null. In particular, we numerically compare the power of the AR and the CLR tests in that scenario. Our numerical results suggest that the AR test outperforms CLR test when the degree of endogeneity is low, however, the order of power performance is reversed when the degree of endogeneity is high. In addition, we propose a new robust test, the Conditional Lagrange Multiplier (CLM) test, which is asymptotically efficient under Strong-IV and exhibits the same power as the AR test under Weak-IV and when alternatives are determined by arbitrarily large deviations from the null.

### 4.5.1 Popular Weak-IV Robust Tests

#### Anderson-Rubin (AR) test

The rejection rule of the AR test is

$$\text{Reject } H_0 \text{ when } AR_n=S_n^{*\prime}S_n^{*}>\chi^2_{l_1,1-\alpha}.$$

By (4.8) and (4.9), under $H_1 : \Delta \to \infty$, $AR_n \sim_a \chi^2_{l_1}(\|\lambda\|^2)$, where $\|\lambda\|^2 = \frac{C'Q_{1\cdot 2}C}{\sigma_v^2}$, $\chi^2_{l_1,1-\alpha}$ is $(1-\alpha)$th quantile of a centred chi-square distribution with the degree of freedom $l_1$, and $\chi^2_{l_1}(\|\lambda\|^2)$ is a non-central chi-square distribution with the non-centrality parameter equal to $\|\lambda\|^2$ and the degree of freedom $l_1$.

**Kleibergen's K (KLM) test**

Though sometimes refereed to as an LM test , this test statistic is different from the usual LM test in that it uses $T_n^*$ (which is asymptotically independent of $S_n^*$) instead of the usual $\hat{\pi}_{1,n}$ to estimate $\pi_{1,n}$. The rejection rule of the KLM test is

$$\text{Reject } H_0 \text{ when } KLM_n = (T_n^{*'} S_n^*)^2 / T_n^{*'} T_n^* > \chi^2_{1,1-\alpha}.$$

Note that the denominator and numerator of KLM are asymptotically independent. Under $H_1 : \Delta \to +\infty$, $KLM_n \sim_a \chi^2_1(C_{KLM})$. The non-centrality parameter $C_{KLM}$ can be described as

$$C_{KLM} = \frac{\left(\left(\mathcal{N} + \frac{\rho}{\sqrt{1-\rho^2}}\lambda\right)' \lambda\right)^2}{\left(\mathcal{N} + \frac{\rho}{\sqrt{1-\rho^2}}\lambda\right)' \left(\mathcal{N} + \frac{\rho}{\sqrt{1-\rho^2}}\lambda\right)},$$

where $\mathcal{N} \sim N(0_{l_1 \times 1}, I_{l_1})$. Under $H_1 : \Delta \to -\infty$, the non-centrality parameter becomes

$$C_{KLM} = \frac{\left(\left(\mathcal{N} - \frac{\rho}{\sqrt{1-\rho^2}}\lambda\right)' \lambda\right)^2}{\left(\mathcal{N} - \frac{\rho}{\sqrt{1-\rho^2}}\lambda\right)' \left(\mathcal{N} - \frac{\rho}{\sqrt{1-\rho^2}}\lambda\right)}.$$

We will numerically evaluate the performance of KLM test in Section 4.5.3.

**Conditional Likelihood Ratio (CLR) test**

The CLR test was proposed by Moreira (2003), and was recommended by Andrews, Moreira and Stock (2006) because it numerically attains the power envelope obtained by the approach of weighted average power. The test statistic can be written as

$$LR_n = S_n^{*\prime} S_n^* - \lambda_{min},$$

where $\lambda_{min}$ is the smallest eigenvalue of $Q_n$. Since we focus on the model with a single endogenous regressor, $LR_n$ can be re-written as

$$LR_n = \frac{1}{2}\left\{ S_n^{*\prime} S_n^* - T_n^{*\prime} T_n^* + \sqrt{\left(S_n^{*\prime} S_n^* - T_n^{*\prime} T_n^*\right)^2 + 4\left(S_n^{*\prime} T_n^*\right)^2} \right\}.$$

The critical value is the $(1-\alpha)$th quantile of the conditional null distribution of $LR_n$ given $T_n^*$, which can be easily computed by the Monte Carlo method. Denote such a critical value as $\kappa_{LR,\alpha}(t)$, for given $T_n^* = t$. The rejection rule of the CLR test is

$$\text{Reject } H_0 \text{ when } LR_n > \kappa_{LR,\alpha}(t).$$

The analytical form of the asymptotic power of CLR test is difficult to obtain, we will numerically evaluate the performance of CLR test in Section 4.5.3.

### 4.5.2 A New Test: Conditional Lagrange Multiplier (CLM) Test

The classical Lagrange Multiplier (LM) test has a distorted size in the Weak-IV scenario. Such a distortion is due to the OLS estimator $\hat{\pi}_{1,n}$. This problem can be solved by computing critical values conditional on $T_n^*$. In this section, we construct the LM test statistic using the OLS estimator $\hat{\pi}_{1,n}$, and compute the critical value conditional on $T_n^*$. This gives rise to our Conditional Lagrange Multiplier (CLM) test. Such a construction ensures that the CLM test is Weak-IV robust. Here we further analyse the power property of CLM test in both Weak-IV and Strong-IV scenarios.

Let us first construct the LM statistic. Note that the asymptotic variance of $\sqrt{n}\left(\frac{Z_1' M_2 Z_1}{n}\pi_{1,n}\right)' \left(\mathrm{AsyVar}(\sqrt{n}S_n)\right)^{-1} S_n$ is

$$\pi_{1,n}' Q_{1\cdot 2} \left(\mathrm{AsyVar}(\sqrt{n}S_n)\right)^{-1} Q_{1\cdot 2}\pi_{1,n},$$

where $\mathrm{AsyVar}(\sqrt{n}S_n)$ is the asymptotic variance of $\sqrt{n}S_n$, which can be consistently estimated. Taking this into account, we construct the LM statistic as

$$LM_n = \frac{\left(\sqrt{n}\left(\frac{Z_1' M_2 Z_1}{n}\hat{\pi}_{1,n}\right)' \left(\mathrm{AsyVar}(\sqrt{n}S_n)\right)^{-1} S_n\right)^2}{\hat{\pi}_{1,n}' Q_{1\cdot 2} \left(\mathrm{AsyVar}(\sqrt{n}S_n)\right)^{-1} Q_{1\cdot 2}\hat{\pi}_{1,n}}$$

$$= \frac{\left(\left(\frac{Z_1' M_2 Z_1}{n}\hat{\pi}_{1,n}\right)' \left(\mathrm{AsyVar}(\sqrt{n}S_n)\right)^{-1/2} S_n^*\right)^2}{\hat{\pi}_{1,n}' Q_{1\cdot 2} \left(\mathrm{AsyVar}(\sqrt{n}S_n)\right)^{-1} Q_{1\cdot 2}\hat{\pi}_{1,n}}, \tag{4.16}$$

where the second line comes from the normalization

$$S_n^* = \left(\mathrm{AsyVar}(\sqrt{n}S_n)\right)^{-1/2} \sqrt{n}S_n.$$

By Assumption 2 and using $\frac{Z_1' M_2 Z_1}{n} \to_p Q_{1\cdot 2}$, we have

$$LM_n - LM \to_p 0, \tag{4.17}$$

where

$$LM = \frac{\left(\hat{\pi}_{1,n}' Q_{1\cdot 2}^{1/2} S_n^*\right)^2}{\hat{\pi}_{1,n}' Q_{1\cdot 2}\hat{\pi}_{1,n}}.$$

Recall that the relation between $\hat{\pi}_{1,n}$ and $T_n^*$ is

$$\begin{aligned} \sqrt{n}\hat{\pi}_{1,n} &= \mathrm{AsyCov}(\sqrt{n}\hat{\pi}_{1,n}, \sqrt{n}S_n)\left(\mathrm{AsyVar}(\sqrt{n}S_n)\right)^{-1/2} S_n^* \\ &\quad + \left(\mathrm{AsyVar}(\sqrt{n}T_n)\right)^{1/2} T_n^*, \end{aligned} \tag{4.18}$$

where $\mathrm{AsyCov}(\sqrt{n}\hat{\pi}_{1,n}, \sqrt{n}S_n)$ is the asymptotic covariance between $\sqrt{n}\hat{\pi}_{1,n}$ and $\sqrt{n}S_n$.

The critical values of the CLM test can be simulated conditional on $T_n^*$ and under $H_0$, similar to those of the CLR test. To simulate critical values, let $R$ be the number of simulation. First we generate $S_{0r}^* \sim N(0, I_{l_2})$ for $r = 1, \ldots, R$. Next, given $T_n^* = t$, we construct $\hat{\pi}_{1,n}$ according to (4.18) with $S_n^*$ replaced by $S_{0r}^*$, and denote the result as $\hat{\pi}_{1,n,r}(t)$ for each $r = 1, \ldots, R$. Then we further compute

$$LM_r^* = \frac{\left(\left(\frac{Z_1' M_2 Z_1}{n}\hat{\pi}_{1,n,r}'(t)\right)' \left(\mathrm{AsyVar}(\sqrt{n}S_n)\right)^{-1/2} S_{0r}^*\right)^2}{\hat{\pi}_{1,n,r}'(t) Q_{1\cdot 2} \left(\mathrm{AsyVar}(\sqrt{n}S_n)\right)^{-1} Q_{1\cdot 2}\hat{\pi}_{1,n,r}(t)},$$

for each $r = 1, \ldots, R$. The critical value for the CLM test, $\kappa_{LM,\alpha}(t)$, is given by the $(1-\alpha)$th empirical quantile of $\{LM_r^* : r = 1, \ldots, R\}$. The rejection rule of the CLM test is

$$\text{Reject } H_0 \text{ when } LM_n > \kappa_{LM,\alpha}(t). \tag{4.19}$$

Now we examine the power of the CLM test under Weak-IV for the alternative $H_1 : \Delta \to +\infty$. Under Weak-IV, (4.18) becomes

$$\sqrt{n}\hat{\pi}_{1,n} = \frac{\sigma_{uv} + \sigma_v^2\Delta}{\sigma(\Delta)} Q_{1\cdot 2}^{-1/2} S_n^* + \frac{\sigma_u^2\sigma_v^2 - \sigma_{uv}^2}{\sigma(\Delta)} Q_{1\cdot 2}^{-1/2} T_n^*.$$

Consequently, when $\Delta \to +\infty$, we obtain

$$\sqrt{n}\hat{\pi}_{1,n} - \sigma_v Q_{1\cdot 2}^{-1/2} S_n^* \to_p 0,$$

and

$$\sqrt{n}\hat{\pi}_{1,n,r}(t) - \sigma_v Q_{1\cdot 2}^{-1/2} S_{0r}^* \to_p 0.$$

Hence,

$$LM_n - S_n^{*'} S_n^* \to_p 0,$$

and

$$LM_r^* - S_{0r}^{*'} S_{0r}^* \to_p 0.$$

The same result can be obtained $H_1 : \Delta \to -\infty$. Therefore, the power of CLM test is the same as that of AR test for arbitrarily large deviations from $H_0$. However, CLM test is more efficient than AR test in the Strong-IV scenario, as we show below.

**Assumption 3. (Strong-IV)** $\pi_{1,n} = \pi_1$, where $\pi_1 \in \mathbb{R}^{l_1}$, and $\pi_1 \neq 0$ is fixed.

**Proposition 3.** *In the linear IV model (4.1) and (4.2), suppose that Assumption 1 and 5 hold. Then the CLM test given by (4.19) is asymptotically efficient against the local alternative* $\gamma = \gamma_0 + \Delta/\sqrt{n}$.

*Proof.* We first compute an effective power upper bound for level $\alpha$ test in the Strong-IV scenario, following the approach of Choi and Schick (2006). Consider a local perturbation $\pi_{1,n} = \pi_1 + \tau/\sqrt{n}$ and a joint test of

$$H_0 : \tau = 0 \text{ and } \gamma = 0,$$

against

$$H_0 : \tau \neq 0 \text{ and } \gamma \neq 0.$$

The asymptotic experiment in the Strong-IV scenario becomes

$$\sqrt{n}\begin{bmatrix} S_n \\ T_n - \pi_1 \end{bmatrix} \sim_a N\left(\begin{bmatrix} \Delta Q_{1\cdot 2}\pi_1 \\ \tau - \frac{\Delta\sigma_{uv}\pi_1}{\sigma_u^2} \end{bmatrix}, \begin{bmatrix} \sigma_u^2 Q_{1\cdot 2} & 0_{l_1\times l_1} \\ 0_{l_1\times l_1} & \frac{\sigma_u^2\sigma_v^2 - \sigma_{uv}^2}{\sigma_u^2} Q_{1\cdot 2}^{-1} \end{bmatrix}\right).$$

Then the asymptotically most powerful unbiased (AMPU) test must be based on the likelihood ratio $LR_n$,

$$\begin{aligned} LR_n &= \frac{\Delta^2 \pi_1' \left(\sqrt{n} S_n\right)}{\sigma_u^2} \\ &+ \left(\tau - \frac{\Delta \sigma_{uv} \pi_1}{\sigma_u^2}\right)' \left(\frac{\sigma_u^2 \sigma_v^2 - \sigma_{uv}^2}{\sigma_u^2} Q_{1\cdot 2}^{-1}\right)^{-1} \sqrt{n}\left(T_n - \pi_1\right) \\ &\sim_a N\left(\text{AsyVar}(LR_n), \text{AsyVar}(LR_n)\right), \text{ where} \\ \text{AsyVar}(LR_n) &= \frac{\Delta^2 \pi_1' Q_{1\cdot 2} \pi_1}{\sigma_u^2} \\ &+ \left(\tau - \frac{\Delta \sigma_{uv} \pi_1}{\sigma_u^2}\right)' \left(\frac{\sigma_u^2 \sigma_v^2 - \sigma_{uv}^2}{\sigma_u^2} Q_{1\cdot 2}^{-1}\right)^{-1} \left(\tau - \frac{\Delta \sigma_{uv} \pi_1}{\sigma_u^2}\right). \end{aligned}$$

Thus, for given $\Delta$ and $\tau$, the AMPU level $\alpha$ test is

$$\text{Reject } H_0 \text{ when } LR_n^2 / \text{AsyVar}(LR_n) > \chi^2_{1,1-\alpha}.$$

The power of the AMPU test is characterized by the non-centrality parameter AsyVar($LR_n$), since

$$LR_n^2 / \text{AsyVar}(LR_n) \sim_a \chi_1^2\left(\text{AsyVar}(LR_n)\right).$$

The power minimizing direction (in terms of distribution to $\pi_1$) is $\tau = \frac{\Delta \sigma_{uv} \pi_1}{\sigma_u^2}$, which leads to the non-centrality parameter equal to $\frac{\Delta^2 \pi_1' Q_{1\cdot 2} \pi_1}{\sigma_u^2}$, for the effective power upper bound in the worst scenario with respect to $\tau$.

Now to consider the power of the CLM test under Strong-IV, note that in the Strong-IV scenario, $\hat{\pi}_{1,n} \to_p \pi_1$. Hence by (4.17),

$$LM_n \to_p \frac{\left(\pi_1' Q_{1\cdot 2}^{1/2} S_n^*\right)^2}{\pi_1' Q_{1\cdot 2} \pi_1} \sim_a \chi_1^2\left(\frac{\Delta^2 \pi_1' Q_{1\cdot 2} \pi_1}{\sigma_u^2}\right),$$

and

$$LM_r^* \sim_a \chi_1^2(0).$$

Therefore, the local power of the CLM test in the Strong-IV scenario is $\Pr\left(V > \chi^2_{1,1-\alpha}\right)$, where $V$ has a non-central chi-square distribution with the degree of freedom 1 and the non-centrality parameter $\frac{\Delta^2 \pi_1' Q_{1\cdot 2} \pi_1}{\sigma_u^2}$. Thus, CLM test attains the effective power bound for a level $\alpha$ test under Strong-IV. □

### 4.5.3 Power Calculations

In this section, we numerically compare the asymptotic power of the AR, KLM, CLR and CLM tests under Weak-IV and when the alternatives are determined by arbitrarily large deviations from the null. Previously we have shown that in this scenario the CLM test has the same asymptotic power as the AR test. From asymptotic experiments (4.8) and (4.9), the distribution of $[S^{*'}, T^{*'}]'$ is summarized by three parameters: the number of instruments $l_1$, the norm of the nuisance parameter $\lambda$ that determines the strength of instruments, and the parameter $\rho$ that measures the degree of endogeneity.

The following Figure 4.1 and Figure 4.2 depict the power of AR (CLM), KLM, CLR tests as well as the power envelope derived in Section 4.3 (denoted as PO) and the power envelope derived in Section 4.4 (denoted as B-B), for testing $H_0 : \Delta = 0$ against $H_1 : \Delta \to +\infty$. Figure 4.1 is for the model with 2 instruments and Figure 4.2 is for the model with 5 instruments. In each graph, the horizontal axis is the magnitude of $||\lambda||$ and the vertical axis is the rejection rate.

Figure 4.1 and Figure 4.2 suggest that (i) in the Weak-IV scenario and for arbitrarily large deviations, none of the AR, KLM, CLR and CLM tests attains the power envelopes (either PO or BB). (ii) The KLM test is dominated by CLR test in those numerical comparisons. (iii) The relative performance of the AR and CLR tests depends on the degree of endogeniety as measured by $|\rho|$. Both (i) and (iii) are quite different from the findings of Andrews, Moreira and Stock (2006), which used a weighted average power as the optimizing criterion and found that CLR test not only numerically dominates KLM and AR tests, but also attains the power envelope given by the optimal invariant and similar test.

Figure 4.3 to Figure 4.6 provide a closer look at the relative power properties of the AR and CLR tests. The horizontal axis is again the magnitude of $||\lambda||$ while the vertical axis is the rejection rate of the AR test minus that of the CLR test. We clearly see that when the degree of endogeneity $|\rho|$ is relatively low, the AR test has a larger power than the CLR test. However, when the degree of endogeneity $|\rho|$ is relatively high, the CLR test has a larger power than the AR test. Moreover, the power of the AR test is more likely to be larger than that of the CLR test as the number of instruments increases.

Figure 4.1: Power comparisons among robust tests, 2 IVs

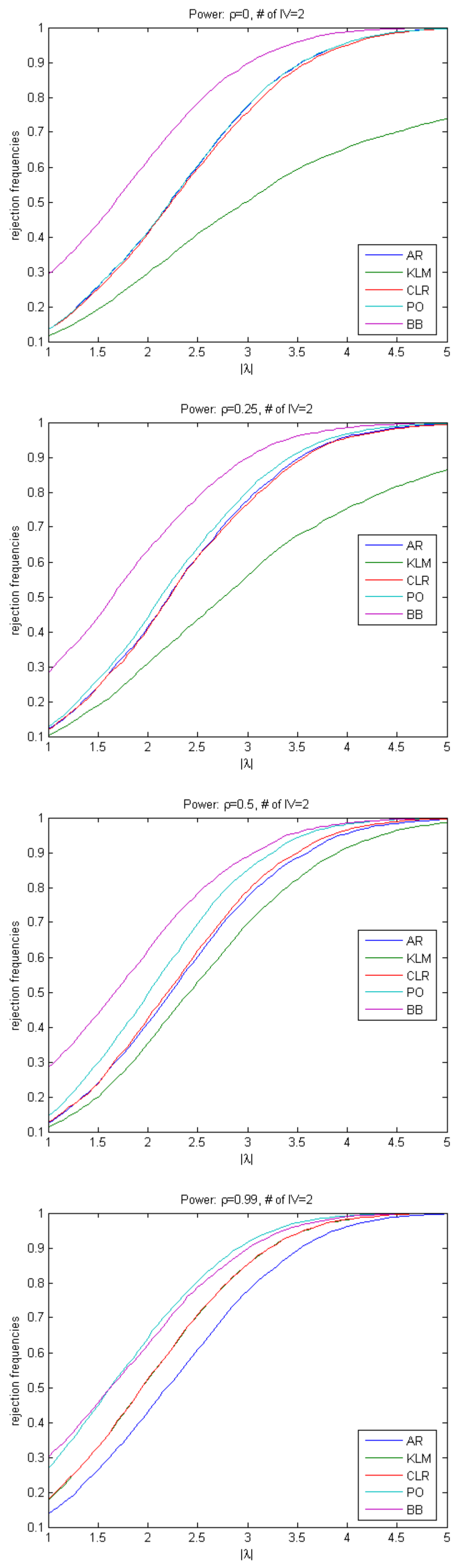


Figure 4.2: Power comparisons among robust tests, 5 IVs

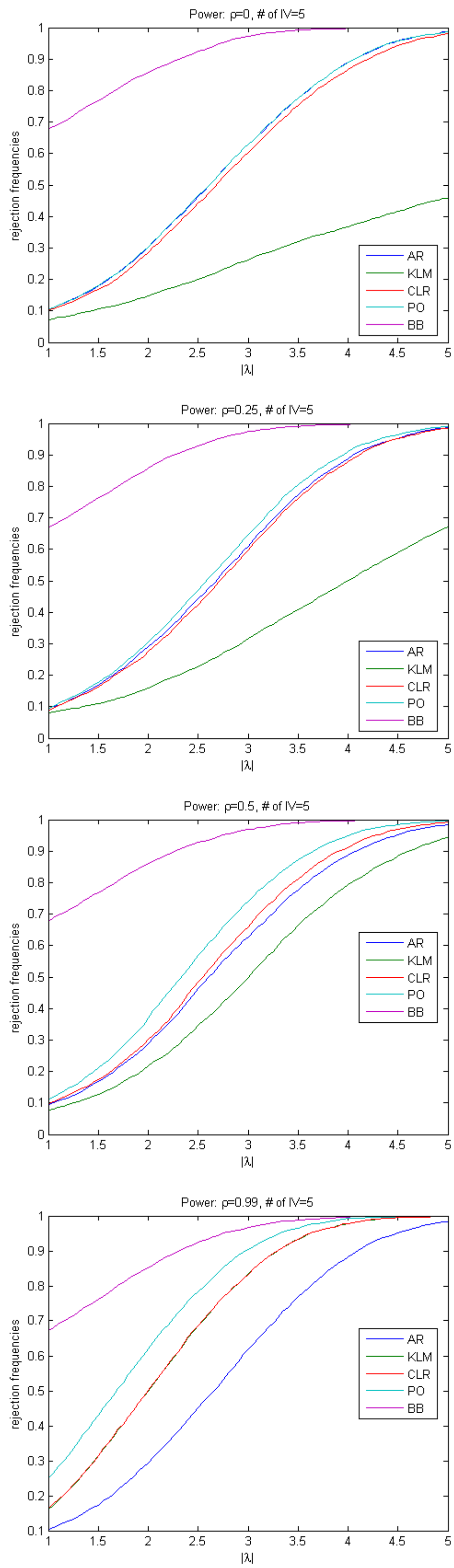


Figure 4.3: Relative power of the AR versus CLR tests, 2 IVs

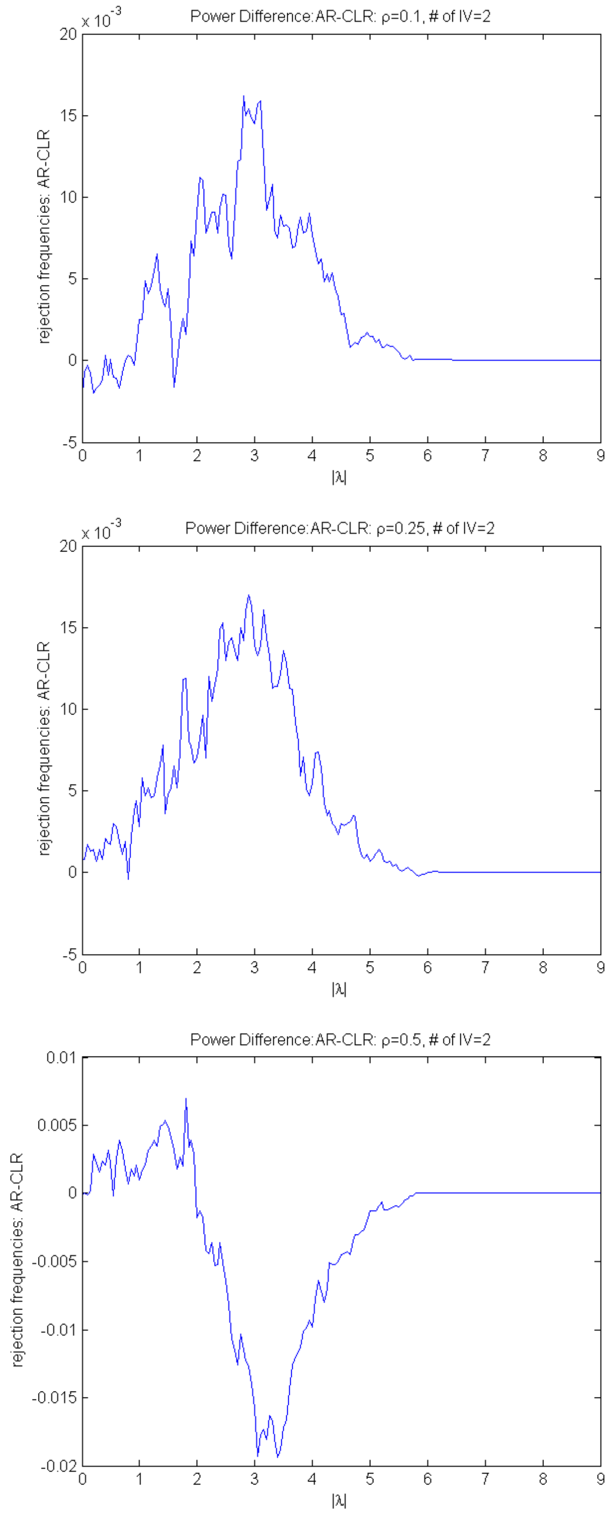


Figure 4.4: Relative power of the AR versus CLR tests, 5 IVs

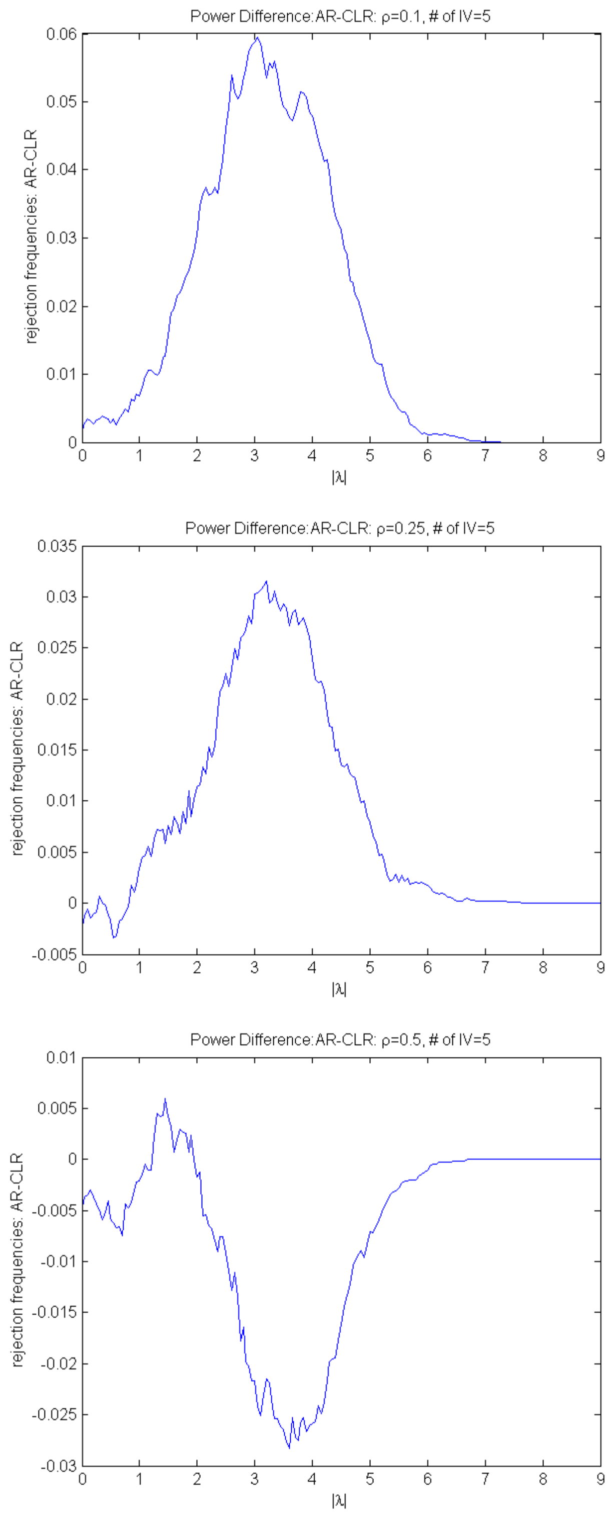


Figure 4.5: Relative power of the AR versus CLR tests, 10 IVs

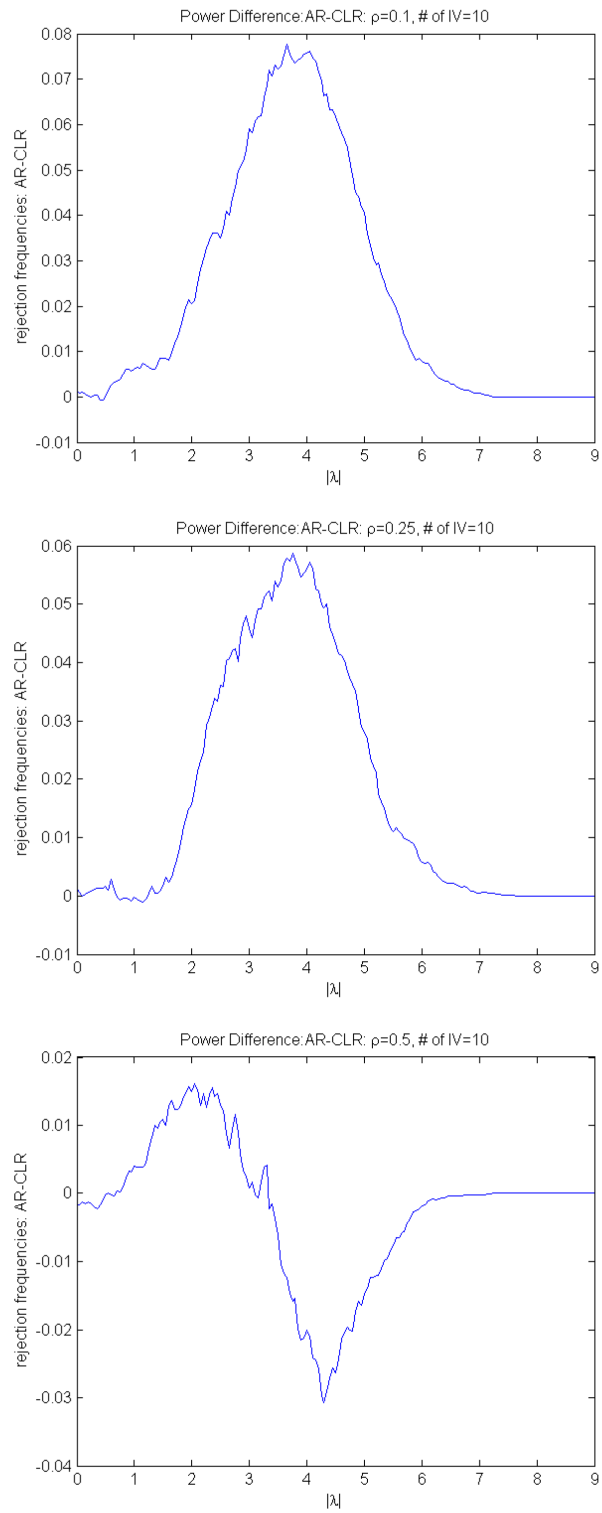
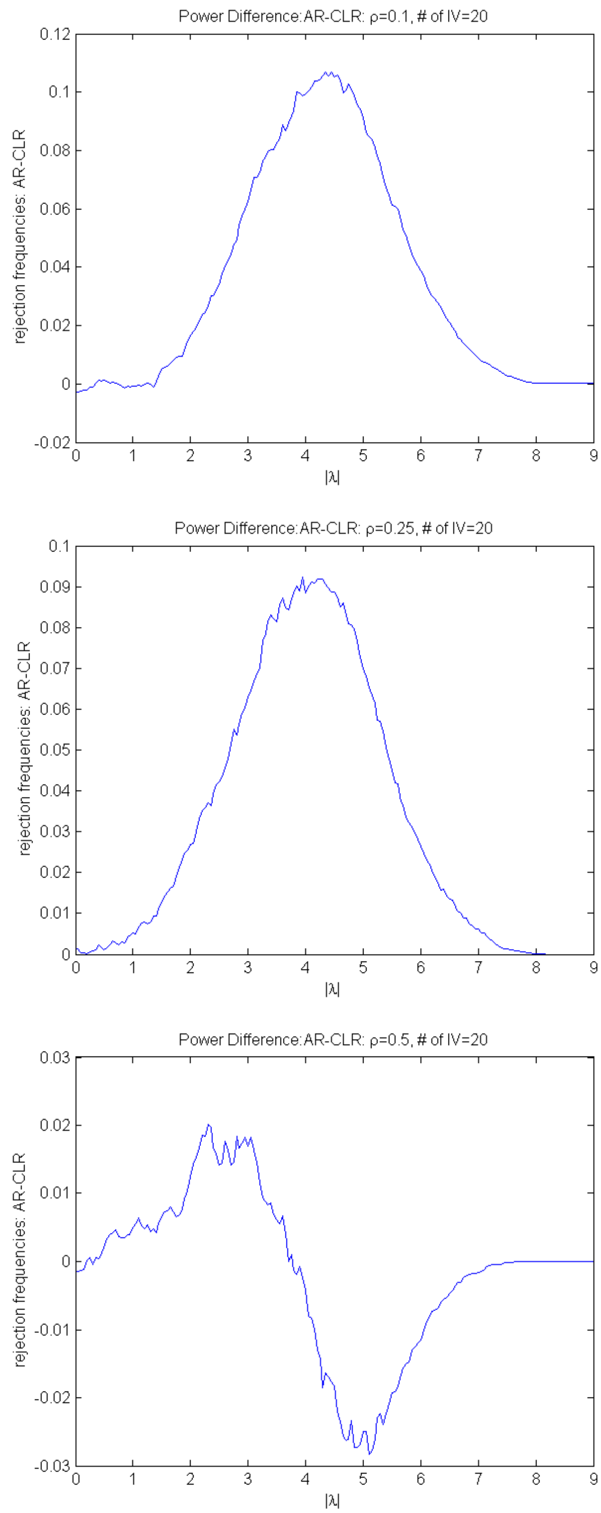


Figure 4.6: Relative power of the AR versus CLR tests, 20 IVs



## 4.6 Heteroskedastic Models

In the previous sections, Assumption 2(c) assumes homoskedasticity for the error terms in model (4.1) and (4.2). This implies the limiting distribution in (4.3). In this section, we allow for heteroskedasticity. We derive the corresponding asymptotic experiments in heteroskedastic models, and then describe the optimal invariant and asymptotically similar test. Also, we analyse the generalized likelihood ratio statistic in heteroskedastic models when alternatives are arbitrarily far away from the null. To facilitate analysis, we apply the following normalization on instruments $Z$ and the nuisance parameters $C$. For $i = 1, ..., n$,

$$\bar{Z}_i = Q_{1\cdot 2}^{-1/2}\tilde{Z}_i, \text{ and } \bar{C} = Q_{1\cdot 2}^{1/2}C,$$

where $\tilde{Z} = M_2 Z_1 = \left(\tilde{Z}_1, \tilde{Z}_2, ..., \tilde{Z}_n\right)'$ and $\bar{Z} = \left(\bar{Z}_1, \bar{Z}_2, ..., \bar{Z}_n\right)'$. Thus we have

$$\frac{1}{\sqrt{n}}\begin{bmatrix} \bar{Z}'u \\ \bar{Z}'v \end{bmatrix} \to_d N\left(0, \begin{bmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{uv} & \Omega_{vv} \end{bmatrix}\right), \tag{4.20}$$

where

$$\begin{aligned}
\Omega_{uu} &= \mathrm{E}[u_i^2\bar{Z}_i\bar{Z}_i'] &&= Q_{1\cdot 2}^{-1/2}\mathrm{E}\left[u_i^2\tilde{Z}_i\tilde{Z}_i'\right]Q_{1\cdot 2}^{-1/2}, \\
\Omega_{vv} &= \mathrm{E}[v_i^2\bar{Z}_i\bar{Z}_i'] &&= Q_{1\cdot 2}^{-1/2}\mathrm{E}\left[v_i^2\tilde{Z}_i\tilde{Z}_i'\right]Q_{1\cdot 2}^{-1/2}, \\
\Omega_{uv} &= \mathrm{E}[u_iv_i\bar{Z}_i\bar{Z}_i'] &&= Q_{1\cdot 2}^{-1/2}\mathrm{E}\left[u_iv_i\tilde{Z}_i\tilde{Z}_i'\right]Q_{1\cdot 2}^{-1/2}.
\end{aligned}$$

Note that the limiting distribution in (4.20) is a characterization of heteroskedasticity.

Similar to the homoskedastic case, we construct $\bar{S}_n$ and $\bar{\pi}_{1,n}$ as

$$\begin{aligned}
\bar{S}_n &= Q_{1\cdot 2}^{-1/2}S_n = && \frac{\bar{Z}'(y_1 - y_2\gamma_0)}{n} = \frac{\bar{Z}'\bar{Z}}{n}\Delta\frac{\bar{C}}{\sqrt{n}} + \frac{\bar{Z}'(u+\Delta v)}{n}, \\
\bar{\pi}_{1,n} &= Q_{1\cdot 2}^{1/2}\hat{\pi}_{1,n} = && \left(\bar{Z}'\bar{Z}\right)^{-1}\bar{Z}'y_2 = \left(\frac{\bar{Z}'\bar{Z}}{n}\right)^{-1}\frac{\bar{Z}'v}{n} + \frac{\bar{C}}{\sqrt{n}},
\end{aligned}$$

where the third equality in each line comes from Assumption 1.

Note that $\mathrm{E}[\bar{Z}_i\bar{Z}_i'] = I_{l_1}$, and using (4.20), we derive the asymptotic distribution of $\bar{S}_n$ and $\bar{\pi}_{1,n}$ as follows,

$$\sqrt{n}\begin{bmatrix} \bar{S}_n \\ \bar{\pi}_{1,n} \end{bmatrix} \sim_a N\left(\begin{bmatrix} \Delta\bar{C} \\ \bar{C} \end{bmatrix}, \begin{bmatrix} \Omega_{uu} + 2\Delta\Omega_{uv} + \Delta^2\Omega_{vv} & \Omega_{uv} + \Delta\Omega_{vv} \\ \Omega_{uv} + \Delta\Omega_{vv} & \Omega_{vv} \end{bmatrix}\right).$$

Let $\bar{T}_n = \sqrt{n}\bar{\pi}_{1,n} - (\Omega_{uv} + \Delta\Omega_{vv}) \left(\Omega_{uu} + 2\Delta\Omega_{uv} + \Delta^2\Omega_{vv}\right)^{-1} \sqrt{n}\bar{S}_n$. We obtain

$$\sqrt{n}\begin{bmatrix} \bar{S}_n \\ \bar{T}_n \end{bmatrix} \sim_a N\left(\begin{bmatrix} \Delta\bar{C} \\ A \end{bmatrix}, \begin{bmatrix} \Omega_{uu} + 2\Delta\Omega_{uv} + \Delta^2\Omega_{vv} & 0_{l_1\times l_1} \\ 0_{l_1\times l_1} & B \end{bmatrix}\right),$$

where

$$A = (\Omega_{uu} + \Delta\Omega_{uv}) \left(\Omega_{uu} + 2\Delta\Omega_{uv} + \Delta^2\Omega_{vv}\right)^{-1} \bar{C},$$

and

$$B = \Omega_{vv} - (\Omega_{uv} + \Delta\Omega_{vv}) \left(\Omega_{uu} + 2\Delta\Omega_{uv} + \Delta^2\Omega_{vv}\right)^{-1} (\Omega_{uv} + \Delta\Omega_{vv}).$$

By letting $S_n^* = \left(\Omega_{uu} + 2\Delta\Omega_{uv} + \Delta^2\Omega_{vv}\right)^{-1/2} \sqrt{n}\bar{S}_n$ and $T_n^* = B^{-1/2}\sqrt{n}\bar{T}_n$, we have

$$\begin{bmatrix} S_n^* \\ T_n^* \end{bmatrix} \sim_a N\left(\begin{bmatrix} \left(\Omega_{uu} + 2\Delta\Omega_{uv} + \Delta^2\Omega_{vv}\right)^{-1/2} \Delta\bar{C} \\ B^{-1/2}A \end{bmatrix}, I_{2l_1\times 2l_1}\right).$$

Define a random vector $\mathcal{N} \sim N(0_{2l_1\times 1}, I_{2l_1})$, and

$$\begin{bmatrix} S^* \\ T^* \end{bmatrix} = \mathcal{N} + \begin{bmatrix} \left(\Omega_{uu} + 2\Delta\Omega_{uv} + \Delta^2\Omega_{vv}\right)^{-1/2} \Delta\bar{C} \\ B^{-1/2}A \end{bmatrix}.$$

Since $\begin{bmatrix} S_n^* \\ T_n^* \end{bmatrix} \to_d \begin{bmatrix} S^* \\ T^* \end{bmatrix}$, the asymptotic experiment for the heteroskedastic model is as follows.

Under $H_0 : \Delta = 0$,

$$\begin{bmatrix} S^* \\ T^* \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ \left(\Omega_{vv} - \Omega_{uv}\Omega_{uu}^{-1}\Omega_{uv}\right)^{-1/2} \bar{C} \end{bmatrix}, I_{2l_1}\right).$$

Under $H_1 : \Delta \to +\infty$, the asymptotic experiment becomes

$$\begin{bmatrix} S^* \\ T^* \end{bmatrix} \sim N\left(\begin{bmatrix} \Omega_{vv}^{-1/2}\bar{C} \\ \left(\Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{uv}\right)^{-1/2} \Omega_{uv}\Omega_{vv}^{-1}\bar{C} \end{bmatrix}, I_{2l_1}\right).$$

Similarly, under $H_1 : \Delta \to -\infty$,

$$\begin{bmatrix} S^* \\ T^* \end{bmatrix} \sim N\left(-\begin{bmatrix} \Omega_{vv}^{-1/2}\bar{C} \\ \left(\Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{uv}\right)^{-1/2} \Omega_{uv}\Omega_{vv}^{-1}\bar{C} \end{bmatrix}, I_{2l_1}\right).$$

The form of the asymptotic experiment for arbitrarily large deviations is determined by the fact that when $\Delta \to +\infty$,

$$\begin{aligned} B^{-1/2}A &= P^{1/2}(\Delta)\left(\frac{\Omega_{uu}}{\Delta}+\Omega_{uv}\right)\left(\frac{\Omega_{uu}+2\Delta\Omega_{uv}+\Delta^2\Omega_{vv}}{\Delta^2}\right)^{-1}\bar{C} \\ &\to \left(\Omega_{uu}-\Omega_{uv}\Omega_{vv}^{-1}\Omega_{uv}\right)^{-1/2}\Omega_{uv}\Omega_{vv}^{-1}\bar{C}, \end{aligned}$$

where

$$\begin{aligned} P(\Delta) &= \left(\frac{\Omega_{uv}}{\Delta}+\Omega_{vv}\right)^{-1}\left(\frac{\Omega_{uu}+2\Delta\Omega_{uv}+\Delta^2\Omega_{vv}}{\Delta^2}\right) \\ &\quad \times \left(\Omega_{uu}-\Omega_{uv}\Omega_{vv}^{-1}\Omega_{uv}\right)^{-1}\left(\frac{\Omega_{uv}}{\Delta}+\Omega_{vv}\right)\Omega_{vv}^{-1}. \end{aligned}$$

Analogous results can be obtained for $\Delta \to -\infty$.

Now let us re-write the mean of the asymptotic statistics $S^*$ and $T^*$. Let

$$D = \left(\Omega_{uu}-\Omega_{uv}\Omega_{vv}^{-1}\Omega_{uv}\right)^{-1/2}\Omega_{uv}\Omega_{vv}^{-1}, \tag{4.21}$$

and note that

$$D'D = \Omega_{vv}^{-1}\Omega_{uv}\left(\Omega_{uu}-\Omega_{uv}\Omega_{vv}^{-1}\Omega_{uv}\right)^{-1}\Omega_{uv}\Omega_{vv}^{-1}.$$

Define

$$\begin{aligned} \begin{bmatrix} \Omega^{vv} & \Omega^{uv} \\ \Omega^{uv} & \Omega^{uu} \end{bmatrix} &= \begin{bmatrix} \Omega_{vv} & \Omega_{uv} \\ \Omega_{uv} & \Omega_{uu} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \Omega_{vv}^{-1}+\Omega_{vv}^{-1}\Omega_{uv}F_u\Omega_{uv}\Omega_{vv}^{-1} & -\Omega_{vv}^{-1}\Omega_{uv}F_u \\ -F_u\Omega_{uv}\Omega_{vv}^{-1} & F_u \end{bmatrix} \\ &= \begin{bmatrix} F_v & -\Omega_{vv}^{-1}\Omega_{uv}F_u \\ -F_u\Omega_{uv}\Omega_{vv}^{-1} & F_u \end{bmatrix}, \end{aligned}$$

where

$$\begin{aligned} F_u &= \left(\Omega_{uu}-\Omega_{uv}\Omega_{vv}^{-1}\Omega_{uv}\right)^{-1}, \\ F_v &= \left(\Omega_{vv}-\Omega_{uv}\Omega_{uu}^{-1}\Omega_{uv}\right)^{-1}. \end{aligned} \tag{4.22}$$

Using this notation, the asymptotic experiment can be re-written as

$$\text{Under } H_0\text{: } \Delta{=}0, \quad \begin{bmatrix} S^* \\ T^* \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ F_v^{1/2}\bar{C} \end{bmatrix}, I_{2l_1}\right), \tag{4.23}$$

$$\text{Under } H_1\text{: } \Delta \to +\infty, \quad \begin{bmatrix} S^* \\ T^* \end{bmatrix} \sim N\left( \begin{bmatrix} \Omega_{vv}^{-1/2}\bar{C} \\ D\bar{C} \end{bmatrix}, I_{2l_1} \right), \tag{4.24}$$

$$\text{Under } H_1\text{: } \Delta \to -\infty, \quad \begin{bmatrix} S^* \\ T^* \end{bmatrix} \sim N\left( \begin{bmatrix} -\Omega_{vv}^{-1/2}\bar{C} \\ -D\bar{C} \end{bmatrix}, I_{2l_1} \right), \tag{4.25}$$

where $D'D = F_v - \Omega_{vv}^{-1}$.

### 4.6.1 The Optimal Rotational Invariant and Asymptotically Similar Test In Heteroskedastic Models

Similar to the Section 4.3, we can describe the power envelope for rotational invariant and asymptotically similar tests in the heteroskedastic case, when alternatives are determined by arbitrarily large deviations. Let

$$\begin{aligned} Q_n &= \begin{bmatrix} Q_{sn} & Q_{stn} \\ Q'_{stn} & Q_{tn} \end{bmatrix} = [S_n^*, T_n^*]'[S_n^*, T_n^*], \\ Q &= \begin{bmatrix} Q_s & Q_{st} \\ Q'_{st} & Q_t \end{bmatrix} = [S^*, T^*]'[S^*, T^*]. \end{aligned}$$

The (infeasible) optimal rotational invariant and asymptotically similar test in this scenario can be established by the following proposition.

**Proposition 4.** *In the linear IV model (4.1) and (4.2), suppose that Assumption 1 and Assumption 2(a),(b),(d) hold, and the heteroskedasticity is characterized by (4.20). Consider a testing problem:*

$$H_0: \ \Delta = 0, \ H_1 : \Delta \to \infty.$$

*Then the test that rejects $H_0$ when*

$$POIS_\infty^H(Q_{sn}, Q_{stn}) = \bar{C}'\Omega_{vv}^{-1}\bar{C}Q_{sn} + 2\bar{C}'\Omega_{vv}^{-1/2}D\bar{C}Q_{stn} > \kappa_\infty^H(Q_{tn}),$$

*maximizes asymptotic power over all asymptotic size $\alpha$ tests as functions of $Q_n$, where $\kappa_\infty^H(Q_{tn})$ is the $(1-\alpha)$th quantile of the null distribution of $POIS_\infty^H(Q_{sn}, Q_{stn})$ conditional on $Q_{tn}$.*

*Proof.* Note that since $Q_n \to_d Q$, we can first restrict our attention on the asymptotic statistic $Q$ and then use the Continuous Mapping Theorem to argue the result for $Q_n$. The random matrix $Q$ has a non-central Wishart distribution. Let $f_Q^0(q_s, q_{st}, q_t)$ and $f_Q^1(q_s, q_{st}, q_t)$ be the densities of $Q$ under $H_0$ and $H_1$ respectively:

$$f_Q^0(q_s, q_{st}, q_t) = K_1 \exp\left( \frac{-\bar{C}'F_v\bar{C}}{2} \right) \det(q)^{\frac{l_1-3}{2}} \exp\left( -\frac{q_s + q_t}{2} \right)$$

$$
\begin{aligned}
& \times(\bar{C}'F_v\bar{C})^{-\frac{l_1-2}{4}} I_{\frac{l_1-2}{2}}\left(\sqrt{\bar{C}'F_v\bar{C}}\right), \\
f_Q^1(q_s, q_{st}, q_t) \quad = \quad & K_1 \text{exp}\left(\frac{-\bar{C}'F_v\bar{C}}{2}\right) \det(q)^{\frac{l_1-3}{2}} \exp\left(-\frac{q_s+q_t}{2}\right) \\
& \times \zeta(q)^{-\frac{l_1-2}{4}} I_{\frac{l_1-2}{2}}\left(\sqrt{\zeta(q)}\right),
\end{aligned}
$$

where

$$
\zeta(q) = \bar{C}'\Omega_{vv}^{-1}\bar{C}q_s + 2\bar{C}'\Omega_{vv}^{-1/2}D\bar{C}q_{st} + \bar{C}'\left(F_v - \Omega_{vv}^{-1}\right)\bar{C}q_t.
$$

Under $H_0$, the random variable $Q_t$ has a non-central chi-square distribution with the degree of freedom $l_1$ and the non-centrality parameter $\bar{C}'F_v\bar{C}$. Under $H_1$, the random variable $Q_t$ has a non-central chi-square distribution with the degree of freedom $l_1$ and the non-centrality parameter $\bar{C}'\left(F_v - \Omega_{vv}^{-1}\right)\bar{C}$. Let $f_{Q_t}^0(q_t)$ and $f_{Q_t}^1(q_t)$ be the densities of $Q_t$ under $H_0$ and $H_1$, respectively.

$$
\begin{aligned}
f_{Q_t}^0(q_t) \quad = \quad & K_2 \text{exp}\left(\frac{-\bar{C}'F_v\bar{C}}{2}\right) q_t^{\frac{l_1-2}{2}} \exp\left(\frac{q_t}{2}\right) \\
& \times\left(\bar{C}'F_v\bar{C}q_t\right)^{-\frac{l_1-2}{4}} I_{\frac{l_1-2}{2}}\left(\sqrt{\bar{C}'F_v\bar{C}q_t}\right), \\
f_{Q_t}^1(q_t) \quad = \quad & K_2 \text{exp}\left(\frac{-\bar{C}'\left(F_v - \Omega_{vv}^{-1}\right)\bar{C}}{2}\right) q_t^{\frac{l_1-2}{2}} \exp\left(\frac{q_t}{2}\right) \\
& \times\left(\bar{C}'\left(F_v - \Omega_{vv}^{-1}\right)\bar{C}q_t\right)^{-\frac{l_1-2}{4}} I_{\frac{l_1-2}{2}}\left(\sqrt{\bar{C}'\left(F_v - \Omega_{vv}^{-1}\right)\bar{C}q_t}\right).
\end{aligned}
$$

Therefore, the likelihood ratio is

$$
\begin{aligned}
LR^H(q) \quad = \quad & \frac{f_Q^1(q_s, q_{st}, q_t)/f_{Q_t}^1(q_t)}{f_Q^0(q_s, q_{st}, q_t)/f_{Q_t}^0(q_t)} \\
= \quad & \frac{\exp\left(-\frac{\bar{C}'\Omega_{vv}^{-1}\bar{C}}{2}\right)\zeta(q)^{-\frac{l_1-2}{4}} I_{\frac{l_1-2}{2}}\left(\sqrt{\zeta(q)}\right)}{\left(\bar{C}'\left(F_v - \Omega_{vv}^{-1}\right)\bar{C}q_t\right)^{-\frac{l_1-2}{4}} I_{\frac{l_1-2}{2}}\left(\sqrt{\bar{C}'\left(F_v - \Omega_{vv}^{-1}\right)\bar{C}q_t}\right)} \\
= \quad & \frac{2^{-\frac{l_1-2}{2}} \exp\left(-\frac{\bar{C}'\Omega_{vv}^{-1}\bar{C}}{2}\right)\sum_{j=0}^{\infty} \frac{(\zeta(q)/4)^j}{j!\Gamma(l_1/2+j)}}{\left(\bar{C}'\left(F_v - \Omega_{vv}^{-1}\right)\bar{C}q_t\right)^{-\frac{l_1-2}{4}} I_{\frac{l_1-2}{2}}\left(\sqrt{\bar{C}'\left(F_v - \Omega_{vv}^{-1}\right)\bar{C}q_t}\right)}.
\end{aligned}
$$

For a given $q_t$, the denominator of $LR^H(q)$ is fixed and the numerator is an increasing function of $\zeta(q)$. Therefore we can construct the conditional test based on $\zeta(q)$ (we can further drop the term $\bar{C}'\left(F_v - \Omega_{vv}^{-1}\right)\bar{C}q_t$, which is constant after fixing $q_t$). The resulting test statistic is

$$POIS_{\infty}^{H}(Q_{sn}, Q_{stn}) = \bar{C}'\Omega_{vv}^{-1}\bar{C}Q_{sn} + 2\bar{C}'\Omega_{vv}^{-1/2}D\bar{C}Q_{stn},$$

and $H_0$ is rejected when $POIS_{\infty}^{H}(Q_{sn}, Q_{stn}) > \kappa_{\infty}^{H}(Q_t)$, where $\kappa_{\infty}^{H}(Q_t)$ is the $(1-\alpha)$th quantile of the distribution of $POIS_{\infty}^{H}(Q_{sn}, Q_{stn})$ conditional on $Q_t$ and under $H_0$. Exactly the same result can be obtained for $H_1 : \Delta \to -\infty$. $\square$

Compared with Proposition 1, we can see that in the homoskedastic case, the optimal test is a function of $Q_n$ and depends only on the parameter $\rho$ that measures the degree of endogeneity. In the heteroskedastic case, the function of $Q_n$ depends on $\bar{C}$, $\Omega_{vv}$ and $F_v$. The expressions for $\Omega_{vv}$ and $F_v$ can be found in (4.20) and (4.22). While the matrix $\Omega_{vv}$ can be estimated consistently, $F_v$ and $\bar{C}$ cannot.

### 4.6.2 The Generalized Likelihood Ratio (GLR) Statistic In Heteroskedastic Models

In this section, we consider the generalized likelihood ratio (GLR) statistic under heteroskedasticity in the Weak-IV scenario and when alternatives are determined by arbitrarily large deviations from the null. We find that the GLR statistic degenerates to AR statistic in this case. Consider the testing problem $H_0 : \Delta = 0$ against $H_1 : \Delta \to +\infty$, based on the asymptotic experiments (4.23) and (4.24), the GLR statistic in heteroskedastic models can be written as

$$\begin{aligned} GLR_n \quad &= \quad S_n^{*\prime}S_n^* \\ &- \quad \min_{\bar{C},\Omega_{uu},\Omega_{uv}}[(S_n^* - \mu_S)'(S_n^* - \mu_S) + (T_n^* - \mu_T)'(T_n^* - \mu_T)], \\ &\text{s.t.} \quad \mu_S = \Omega_{vv}^{-1/2}\bar{C}, \mu_T = D\bar{C}. \end{aligned} \tag{4.26}$$

Recall that $\Omega_{vv}$ can be treated as known (since it can be estimated consistently), and the matrix $D$ defined in (4.21) depends on $\Omega_{uu}$ and $\Omega_{uv}$, which are unknown. The next proposition states that the GLR statistic under heteroskedasticity reduces to the AR statistic in the Weak-IV scenario and when alternatives are determined by arbitrarily large deviations from the null.

**Proposition 5.** *In the linear IV model (4.1) and (4.2), suppose that Assumption 1 and Assumption 2(a),(b),(d) hold, and the heteroskedasticity is characterized by (4.20). Consider a testing problem*

$$H_0: \ \Delta = 0, \ H_1 : \Delta \to +\infty.$$

*Then the GLR statistic defined by (4.26) equals to the AR statistic $S_n^{*\prime} S_n^*$.*

*Proof.* Since for any value $s^*$ of $S_n^*$, we can always set the minimizer $\bar{C} = \Omega_{vv}^{1/2} s^*$, it suffices to show that for an arbitrary value[27] $s^*$ of $S_n^*$ and $t^*$ of $T_n^*$, there exist some $\Omega_{uu}^*$ and $\Omega_{uv}^*$ such that $t^* = Rs^*$ and $\begin{bmatrix} \Omega_{uu}^* & \Omega_{uv}^* \\ \Omega_{uv}^* & \Omega_{vv} \end{bmatrix}$ is positive semi-definite (recall that $\Omega_{vv}$ is known), where

$$R = D\Omega_{vv}^{1/2} = \left(\Omega_{uu}^* - \Omega_{uv}^* \Omega_{vv}^{-1} \Omega_{uv}^*\right)^{-1/2} \Omega_{uv}^* \Omega_{vv}^{-1/2}. \tag{4.27}$$

Let $t^* = \left(t_1^*, ..., t_{l_1}^*\right)'$ and $R = (R_1, ..., R_{l_1})'$. Note that $t^* = Rs^*$ is equivalent to $t_k^* = R_k' s^*$ for $k = 1, ..., l_1$. Therefore, $R$ is determined by

$$R_1 = (t_1^*/s_1^*, 0, ..., 0)', R_2 = (0, t_2^*/s_2^*, ..., 0)', ..., R_1 = \left(0, ..., t_{l_1}^*/s_{l_1}^*\right)',$$

and

$$R = diag[t_1^*/s_1^*, t_2^*/s_2^*, ..., t_{l_1}^*/s_{l_1}^*]. \tag{4.28}$$

The next step is to show that the $R$ satisfying (4.28) can be constructed using some $\Omega_{uu}^*$ and $\Omega_{uv}^*$ such that the resulting matrix $\begin{bmatrix} \Omega_{uu}^* & \Omega_{uv}^* \\ \Omega_{uv}^* & \Omega_{vv} \end{bmatrix}$ is positive semi-definite. By (4.27) and $D'D = F_v - \Omega_{vv}^{-1}$, we have

$$\begin{aligned} R'R &= \Omega_{vv}^{1/2} F_v \Omega_{vv}^{1/2} - I \\ &= \Omega_{vv}^{1/2} \left(\Omega_{uu}^* - \Omega_{uv}^* \Omega_{vv}^{-1} \Omega_{uv}^*\right)^{-1} \Omega_{vv}^{1/2} - I. \end{aligned}$$

Rearranging the terms leads to

$$\Omega_{uu}^* = \Omega_{uv}^* \Omega_{vv}^{-1} \Omega_{uv}^* + \Omega_{vv}^{1/2} \left(R'R + I\right)^{-1} \Omega_{vv}^{1/2}. \tag{4.29}$$

This means that for an arbitrary $\Omega_{uv}^* \in \mathbb{R}^{l_1} \times \mathbb{R}^{l_1}$, we can construct an $\Omega_{uu}^*$ as in (4.29). By construction, the resulting $\Omega_{uv}^*$ and $\Omega_{uu}^*$ satisfy $t^* = Rs^*$.

[27] except for $s^* = 0$. Note that we have $\lim_{n \to +\infty} \Pr(S_n^* = 0) = 0$.

To see that $\begin{bmatrix} \Omega^*_{uu} & \Omega^*_{uv} \\ \Omega^*_{uv} & \Omega_{vv} \end{bmatrix}$ is positive semi-definite, take an arbitrary vector $c = (c_1', c_2')'$, where $c_j \in \mathbb{R}^{l_1}$ for $j = 1, 2$, and $c \neq 0$:

$$
\begin{aligned}
& \begin{bmatrix} c_1' & c_2' \end{bmatrix} \begin{bmatrix} \Omega^*_{uu} & \Omega^*_{uv} \\ \Omega^*_{uv} & \Omega_{vv} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \\
= {} & c_1'\Omega^*_{uv}\Omega_{vv}^{-1}\Omega^*_{uv}c_1 + 2c_1'\Omega^*_{uv}c_2 + c_2'\Omega_{vv}c_2 + c_1'\Omega_{vv}^{1/2}\left(R'R + I\right)^{-1}\Omega_{vv}^{1/2}c_1 \\
= {} & \left(\Omega_{vv}^{-1/2}\Omega^*_{uv}c_1 + \Omega_{vv}^{1/2}c_2\right)'\left(\Omega_{vv}^{-1/2}\Omega^*_{uv}c_1 + \Omega_{vv}^{1/2}c_2\right) \\
& + c_1'\Omega_{vv}^{1/2}\left(R'R + I\right)^{-1}\Omega_{vv}^{1/2}c_1 \geq 0.
\end{aligned}
$$

The first equality comes from (4.29). The last inequality comes from the facts that the first term on the right hand side is $||\Omega_{vv}^{-1/2}\Omega^*_{uv}c_1 + \Omega_{vv}^{1/2}c_2||^2$ and $R'R + I$ is positive semi-definite. □

The GLR statistic is calculated by maximizing the likelihood ratio with respect to nuisance parameters unrestricted by $H_0$ or $H_1$, subject to the restrictions imposed by relationships between the means of the asymptotic experiment. Under the null, relationships between the asymptotic means of $S_n^*$ and $T_n^*$ do not provide any restriction on the nuisance parameters. Therefore, after the maximization, the likelihood under the null becomes the AR statistic. Under the alternative hypothesis $H_1 : \Delta \to +\infty$, the asymptotic means of $S_n^*$ and $T_n^*$ must satisfy certain relationships, which are very different in homoskedastic and heteroskedastic models. In the homoskedastic case, under $H_1$ the asymptotic means of $S_n^*$ and $T_n^*$ are proportional (related by a scaling factor $\frac{\rho}{\sqrt{1-\rho^2}}$, see (4.8)). This proportional restriction determines the CLR statistic. In the heteroskedastic case, relationships between the asymptotic means of $S_n^*$ and $T_n^*$ are characterized by the matrix $R$ in (4.27) which depends on the unknown matrices $\Omega_{uu}$ and $\Omega_{uv}$. However, as shown in the proof of Proposition 5, the structure of $R$ is general enough, so that the unknown matrices $\Omega_{uu}$ and $\Omega_{uv}$ can produce any relationship between the asymptotic means of $S_n^*$ and $T_n^*$. In other words, relationships between the asymptotic means of $S_n^*$ and $T_n^*$ are non-restrictive. As a result, in the heteroskedastic case, the maximized likelihood ratio only contains the term obtained from the likelihood under the null, which is the AR statistic.

## 4.7 Concluding Remarks

In this chapter, we consider efficient inference for the coefficient on the endogenous variable in linear regression models with weak instruments. We focus on the power of tests when alternatives are determined by arbitrarily large deviations from the null. We derive the power envelope for such alternatives in the Weak-IV scenario. Then we compare the power properties of popular Weak-IV robust tests, focusing on the AR and CLR tests. We find that their relative performance depends on the degree of endogeniety in the model. This is different from Andrews, Moreira and Stock (2006), which found that CLR numerically dominates AR when the weighted average power is concerned. In addition, we propose the CLM test, which is asymptotically efficient in the Strong-IV scenario, robust to Weak-IV, and exhibits the same power as the AR test for arbitrarily large deviations from the null. We also study the heteroskedastic case, and find that the generalized likelihood ratio statistic under heteroskedasticity reduces to the AR statistic in the Weak-IV scenario and when alternatives are determined by arbitrarily large deviations from the null.

Our analysis can also be extended to a more general minimum distance estimation (MDE) framework (see Newey and McFadden (1994) for a treatment of classical minimum distance estimation). Econometric models such as censored regression with endogenous regressors[28] and DSGE models[29] fit into the MDE framework. Magnusson (2010) described the AR, KLM and CLR tests in the context of MDE, however, the question of efficiency remains open. In the MDE framework, one can derive the asymptotic experiments analogous to those in this chapter. Based on that, one can compute the power envelope and compare the power properties of weak-IV robust tests. Moreover, the CLM test proposed in this chapter can also be used in the MDE framework.

[28]See Powell (1984, 1986) for a semi-parametric quantile regression estimator for the reduced-form parameters. Estimators for the structural parameters can be constructed following the approach of Amemiya (1978).

[29]See for example, Hnatkovska, Marmer and Tang (2012).

# Bibliography


[1] Aguirregabiria, V. and P. Mira (2013) Identification of games of incomplete information with multiple equilibria and common unobserved heterogeneity. Working paper

[2] Amemiya, T. (1978) The estimation of a simultaneous equation generalized probit model. *Econometrica* 46 (5), 1193–1205.

[3] Anderson, T. W. and H. Rubin (1949) Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* **20**, 46–63.

[4] Andrews, D. W. K., M. Moreira and J. Stock (2006) Optimal invariant similar tests for instrumental variables regression. *Econometrica* **74**, 715–752.

[5] Aradillas-Lopez, A., (2010) Semiparametric estimation of a simultaneous game with incomplete information. *Journal of Econometrics,* **157**, 409-431.

[6] Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2010). Estimating static models of strategic interactions. *Journal of Business and Economic Statistics*, **28**, 469-482.

[7] Berry, S., and P. A. Halie (2014) Identification in Differentiated Products Markets Using Market Level Data. *Econometrica* forthcoming

[8] Berry, S., and P. A. Halie (2013) Identification in a class of nonparametric simultaneous equations models. Cowles Foundation Discussion Paper No. 1787.

[9] Berry, S., J. Levinsohn and A. Pakes (1999) Voluntary Export Restraints on Automobiles: Evaluating a Trade Policy. *The American Economic Review*, **89**, 400-431.

[10] Bickel, P. and M. Rosenblatt (1973) On some global measures of the deviations of density function estimates. *Annuals of Statistics*, **26**, 1826-1856.

[11] Billingsley, R. (1986) *Probability and Measure*, Wiley, New York

[12] Blumenson, L. E. (1960) A derivation of n-dimensional spherical coordinates. *The American Mathematical Monthly*, **67**, 63-66

[13] Bowman, A. W., A. Pope and B. Ismail (2006) Detecting discontinuities in nonparametric regression curves and surfaces. *Statistics and Computing*, **16**, 377–390.

[14] Borkovsky, R. N., P. Ellickson, B. Gordon, V. Aguirregabiria, P. Gardete, P. Grieco, T. Gureckis, T. Ho, L. Mathevet and A. Sweeting (2014) Multiplicity of equilibria and information structures in empirical games: challenges and prospects. Session at the 9th Triennial Choice Symposium.

[15] Brock, W., and S. N. Durlauf (2001a) Discrete choice with social interactions. *Review of Economic Studies,* **68**, 235-260.

[16] Brock, W., and S. N. Durlauf (2001b) Interaction-based models. in Heckman, J and E. Leamer edited *Handbook of Econometrics*, vol 5. Elsevier Science.

[17] Brock, W., and S. N. Durlauf (2007) Identification of binary choice modles with social interactions. *Journal of Econometrics*, **140**, 52-75.

[18] Cattaneo, M. D., R. Crump and M. Jansson (2012) Optimal inference for instrumental variables regression with non-gaussian errors. *Journal of Econometrics* **167**, 1–15.

[19] Chernozhukov, V., D. Chetverikov and K. Kato (2012) Gaussian approximation of suprema of empirical process. Working paper

[20] Chernozhukov, V., D. Chetverikov and K. Kato (2013) Anti-concentration and honest adaptive confidence bands. Working paper

[21] Chernozhukov, V., S. Lee and A. Rosen (2013) Intersection bounds: estimation and inference. *Econometrica,* **81**, 667-737.

[22] Chu, C. K. and P. E. Cheng (1996) Estimation of jump points and jump values of a density function. *Statistica Sinica* **6**, 79-95.

[23] Choi, S., W. Hall and A. Schick (1996) Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Annals of Statistics* **24**, 841–861.

[24] Dagsvik, J. and B. Jovanovic (1994). Was the Great Depression a low-level equilibrium? *European Economic Review*, **38**, 1711-1729.

[25] Delgado, M. A and J. Hidalgo (2000) Nonparametric inference on structural breaks. *Journal of Econometrics,* **96**, 113–144

[26] Dufour, J.-M (1997) Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica* **65**, 1365–1387.

[27] Echenique, F., and I. Komunjer (2009) Testing models with multiple equilibria by quantile methods. *Econometrica*, **77**, 1281-1297.

[28] Echenique, F., and I. Komunjer (2007) Testing models with multiple equilibria by quantile methods. Caltech HSS working paper 1244R.

[29] Ekeland, I., J. J. Heckman and L. Nesheim (2004) Identification and estimation of hedonic models. *Journal of Political Economy*, **112**, S60-S109.

[30] Gijbels, I. and A. C. Goderniaux (2004) Bootstrap test for change-points in nonparametric regression, *Journal of Nonparametric Statistics,* **16**, 591-611.

[31] Giné, E. and R. Nickl (2010) Confidence bands in density estimation. *Annuals of Statistics*, **38**, 1122-1170.

[32] Hall, P. and D. M. Titterington (1992) Edge-preserving and peak-preserving smoothing. *Technometrics,* **34**, 429-440.

[33] Hnatkovska, V., V. Marmer and Y. Tang, (2012) Comparison of misspecified calibrated models: The minimum distance approach. *Journal of Econometrics* **169**, 131–138.

[34] Horowitz, J. L, (1996) Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica*, **64**, 103-137.

[35] Johnston, G. (1982) Probabilities of maximal deviations for nonparametric regression function estimates. *Journal of Multivariate Analysis*, **12**, 402-414.

[36] Jovanovic, B. (1989) Observable implications of models with multiple equilibria. *Econometrica*, **57**, 1431-1437.

[37] Kleibergen, F (2002) Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica* **70**, 1781–1803.

[38] Kleibergen, F (2007) Generalizing weak instrument robust iv statistics towards multiple parameters, unrestricted covariance matrices and identification statistics. *Journal of Econometrics* **139**, 181–216.

[39] Ksay, M. (2012) Nonparametric inference on the number of equilibria, Working paper.

[40] L. H. Loomis and S. Sternberg. (1968). *Advanced calculus*. Reading, Massachusetts: Addison–Wesley.

[41] Liu, N., Vuong, Q., and Xu, H. (2013) Rationalization and nonparametric identification of discrete games with correlated types. Working paper.

[42] Magnusson, L. (2010) Inference in limited dependent variable models robust to weak identification. *Econometrics Journal* **13**, 56–79.

[43] Matzkin, R. L. (2008) Identification in nonparametric simultaneous equations models. *Econometrica*, **76**, 945-978.

[44] Mills, B., M. Moreira and L. Vilela (2014) Tests based on t-statistics for IV regression with weak instruments. *Journal of Econometrics* **182**, 351–363.

[45] Moreira, M. J. (2001) Tests with correct size when instruments can be arbitrarily weak, unpublished manuscript, Department of Economics, University of California, Berkeley.

[46] Moreira, M. J., (2003) A conditional likelihood ratio test for structural models. *Econometrica* **71**, 1027–1048.

[47] Munkres, J. (1999) *Topology*, Pearson.

[48] Müller, H. G. (1992) Change-points in nonparametric regression analysis. *Annuals of Statistics*, **20**, 737-761.

[49] Müller, H. G. and U. StadtmMüller (1999). Discontinuous versus smooth regression. *Annuals of Statistics*, **27**, 299–337.

[50] Newey, W. K and D. McFadden (1994) Large sample estimation and hypothesis testing. In: Engle, R. F., McFadden, D. L. (Eds.), *Handbook of Econometrics*. Vol. IV. Elsevier, Amsterdam, Ch. 36, pp. 2111–2245.

[51] Pakes, A and D. Pollard (1989) Simulation and the Asymptotics of Optimization Estimators. *Econometrica*, **57**, 1027-1057.

[52] Paula, A., (2012) Econometric analysis of games with multiple equilibria. *Annual Review of Economics*, **5**, 107-131.

[53] Paula, A and X. Tang (2012) Inference of signs of interaction effects in simultaneous games with incomplete information. *Econometrica*, **80**, 143–172.

[54] Powell, J. L.(1984) Least absolute deviations estimation for the censored regression model. *Journal of Econometrics* **25**, 303–325.

[55] Powell, J. L.(1986) Censored regression quantiles. *Journal of Econometrics* **32**, 143–155.

[56] Qiu, P. (1997) Nonparametric estimation of jump surface. *Sankhya: The Indian Journal of Statistics*, **59**, 268-294.

[57] Qiu, P. (2002) A nonparametric procedure to detect jumps in regression surfaces. *Journal of Computational and Graphical Statistics*, **11**, 799-822.

[58] Staiger, D., and J., Stock (1997) Instrumental variables regression with weak instruments. *Econometrica* **65**,

[59] Sweeting, A. (2009) The strategic timing incentives of commercial radio stations: an empirical analysis using multiple equilibria. *Rand Journal of Economics*, **40**, 710-742.

[60] Xiao, R. (2014) Identification and estimation of Incomplete Information Games with multiple equilibria. Working paper

[61] Xu, H and Y. Wan (2014) Semiparametric identification and estimation of binary discrete games of incomplete information with correlated private signals., *Journal of Econometrics*, **182**, 235-246, 2014.

[62] van der Vart, A, W and J. A. Wellner (2000) *Weak Convergence and Empirical Processes: With Applications to Statistics,* Springer.

# Appendix A

# Appendix for Chapter 2

## A.1 Proofs for Section 2.3

**Proof of Lemma 1.**

(i) Suppose to the contrary, there are $y_1, y_2 \in A_m$, and $r(y_1) = r(y_2)$. By $A_m = q_m(B_m)$, there are $v_1, v_2 \in B_m$ such that $y_1 = q_m(v_1), y_2 = q_m(v_2)$. Then by Definition 1(i) and $r(y_1) = r(y_2)$, we have

$$v_1 = r(q_m(v_1)) = r(y_1) = r(y_2) = r(q_m(v_2)) = v_2.$$

(ii) For an arbitrary $y \in \bar{\mathcal{Y}}$, by Definition 1(ii), we either have $y \in q_m(B_m) = A_m$ or $y = q_m(\tilde{v})$, where $\tilde{v} = \lim_{n\to\infty} v_n$, for some sequence $v_n \in B_m$. If it is the first case, the proof is done. If it is the second case, then by continuity of $q_m(\cdot)$, $y = \lim_{n\to\infty} y_n$ for $y_n = q_m(v_n) \in A_m$. That is, $y \in \bar{A}_m$.

(iii) Suppose to the contrary, there exists certain $y_0 \in A_1 \cap A_2 \neq \varnothing$. That is, $y_0 \in q_1(B_1)$ and $y_0 \in q_2(B_2)$. Also, we have $r(y_0) = v_1 \in B_1$ and $r(y_0) = v_2 \in B_2$, which implies $v_1 = v_2 \in B_1 \cap B_2$. But this contradicts Definition 1(iii).

(iv) By (C1), there exists a unique function $q(v)$ such that for any $(y, v) \in \bar{\mathcal{Y}}\times$ int$(\mathcal{V})$ satisfying $r\,(y) = v$, $y = q(v)$. We need to show that $q(v)$ is twice continuously differentiable on int$(\mathcal{V})$. By Assumptions D1 to D3 and applying Implicit Function Theorem (see, for example, Theorem 9.4 in Loomis and Sternberg (1968))) on an arbitrary point $(y_0, v_0) \in \bar{\mathcal{Y}}\times$ int$(\mathcal{V})$, there are neighbourhoods $A$ of $y_0$ and $B$ of $v_0$ on which $r(y) = v$ uniquely defines $y$ as a function of $v$. That is, there is a function $\xi : B \to A$ such that $r(\xi(v)) = v$ for all $v \in B$; for each $v \in B$, $\xi(v)$ is the unique solution to $r(y) = v$ lying in $A$, and $\xi(v)$ is twice continuously differentiable on $B$. Because $(y_0, v_0)$ is arbitrary and $q(v)$ is the only solution to $r(y) = v$ by (C1), we must have $q(v) = \xi(v)$ for all $v \in$ int$(\mathcal{V})$. Thus $q(v)$ is twice continuously differentiable on a neighbourhood of every $v \in$ int$(\mathcal{V})$. This immediately leads to the desired result.

(v) When the model is in (C2) or (C3), there exists some $v_M \in (\mathcal{V})$ such that $r(y_1) = r(y_2) = v_M$ for some $y_1 \neq y_2$. Suppose $M = 1$, there is a function $q : \mathrm{int}(\mathcal{V}) \to \bar{\mathcal{Y}}$ such that $y = q(v)$ for all $(y, v) \in \bar{\mathcal{Y}} \times \mathrm{int}(\mathcal{V})$ satisfying $r(y) = v$. But this means $y_1 = q(v_M)$ and $y_2 = q(v_M)$, which can not happen. **Q.E.D.**

**Proof of Lemma 2.**
Let $\xi \sim U[0,1]$ independent of $(X, U)$. Given $X = x$. We have

$$Y = \sum_{m=1}^{M} 1\left\{\sum_{k=0}^{m-1} \lambda_k(x, U) \leq \xi < \sum_{k=0}^{m} \lambda_k(x, U)\right\} q_m(U + g(x)),$$

where we define $\lambda_0(x, u) = 0$, and $q_m(v) = c$ for some constant, for any $v = g(x) + u \notin B_m$. This does not affect the generation of $Y$ because if $v \notin B_m$, we must have $m \notin \mathcal{M}(v)$ and thus $\lambda_m(x, u) = 0$ for all $(x, u)$. Because $A_{sh}$ are mutually exclusive across $s$ and $h$, for any $y \in \cup_{m=1}^{M} \cup_{l \in L_m} A_{ml}$, there is a unique $(s, h)$ such that $y \in A_{sh}$. For any open set $Q \subset A_{sh}$, observe that

$$\begin{aligned}
& \Pr(Y \in Q | X = x) \\
= \; & \Pr\left(\sum_{m=1}^{M} 1\left\{\sum_{k=0}^{m-1} \lambda_k(x, U) \leq \xi < \sum_{k=0}^{m} \lambda_k(x, U)\right\} q_m(U + g(x)) \in Q | X = x\right) \\
= \; & \Pr\left(\{q_s(U + g(x)) \in Q\} \cap \left\{\sum_{k=0}^{s-1} \lambda_k(x, u) \leq \xi < \sum_{k=0}^{s} \lambda_k(x, u)\right\}\right) \\
= \; & \int\int 1\{q_s(u + g(x)) \in Q\} 1\left\{\sum_{k=0}^{s-1} \lambda_k(x, u) \leq v < \sum_{k=0}^{s} \lambda_k(x, u)\right\} \\
& \times f_{U,\xi}(u, v) du dv \\
= \; & \int_{u \in r(Q) - g(x)} \left[\int 1\left\{\sum_{k=0}^{s-1} \lambda_k(x, u) \leq v < \sum_{k=0}^{s} \lambda_k(x, u)\right\} f_{\xi|U}(v|u) dv\right] \\
& \times f_U(u) du \\
= \; & \int_{u \in r(Q) - g(x)} \lambda_s(x, u) f_U(u) du \\
= \; & \int_{y \in Q} \lambda_s(x, r(y) - g(x)) f_U(r(y) - g(x)) |J(y)| dy.
\end{aligned}$$

The second equality comes from the fact that $q_m(u + g(x)) \notin Q$ for any $m \neq s$ as well as the independence between $(U, \xi)$ and $X$. The fourth equality uses the fact that $q^{-1}(\cdot) = r(\cdot)$ is one to one over $A_s$ (Definition 1(i)), and thus over $Q \subset A_{sh}$. The fifth equality comes from the independence between $U$ and $\xi$ as well as the uniform distribution of $\xi$. The last equality comes from the fact that $r(y) - g(x)$ is one to one in $y$ within $A_s$ (and thus within $Q$), $\lambda_s(\mathcal{M}(A_{sh}), x, u) f_U(u)$ is non-negative, and Theorem 17.2 of Billingsley (1986) p229.

Therefore, for any open set $O \subset \cup_{m=1}^{M} \cup_{l \in L_m} A_{ml}$,

$$\begin{aligned} & \Pr(Y \in O | X = x) \\ = & \sum_{m=1}^{M} \sum_{l \in L_m} \int_{y \in O \cap A_{ml}} \lambda_m(x, r(y) - g(x)) f_U(r(y) - g(x)) |J(y)| dy \\ = & \int_{y \in O} \sum_{m=1}^{M} \sum_{l \in L_m} 1\{y \in A_{ml}\} \lambda_m(x, r(y) - g(x)) f_U(r(y) - g(x)) |J(y)| dy. \end{aligned}$$

The desired result follows. **Q.E.D.**

**Proof of Proposition 1.**

If the model is in (C1), that is, $r(\cdot)$ is one to one, then by Assumptions D2 to D4, we have

$$f_{Y|X}(y|x) = f_U(r(y) - g(x)) |J(y)| \text{ for } x \in \mathcal{X}, y \in \mathrm{int}(\mathcal{Y}(x)),$$

and thus (i) and (ii) follows from Assumption D2 and D4.

If the model is in (C2), given any $(y, x)$, we have a unique $s^*$ such that $\lambda_{s^*}(x, r(y) - g(x)) = 1$ and $\lambda_s(x, r(y) - g(x)) = 0$ for all $s \neq s^*$. (Note that $s^*$ may depend on $y$ and $x$.)

(i) For any $x \in \mathcal{X}$ and $y \in \left(\cup_{l \in L_{s^*}} A_{s^* l}\right) \cap \mathrm{int}(\mathcal{Y}(x))$, we have

$$\lambda_{s^*}(x, r(y) - g(x)) = 1, \tag{A.1}$$

where $s^*$ may depend on $(y, x)$.

Consider any sequence $\{y_n\}$ such that $y_n \in \mathrm{int}(\mathcal{Y}(x))$ and $y_n \to y$, we must have

$$\lambda_{s^*}(x, r(y_n) - g(x)) = 1,$$

for sufficiently large $n$, where $s^*$ is the same as the one in (A.1). Suppose it is not. Then for any $N$, there exists an $n \geq N$ such that $\lambda_{s^*}(x, r(y_n) - g(x)) = 0$. Choose large enough $N$ such that $y_n \in B_\epsilon(y) \subset A_{s^*}$ for arbitrarily small $\epsilon$, where $B_\epsilon(y)$ is an open ball centred in $y$ and with radius $\epsilon$. Together with Lemma 2, we have $f_{Y|X}(y_n|x) = 0$, which contradicts with $y_n \in \mathrm{int}(\mathcal{Y}(x))$. Therefore, by Lemma 2, we have

$$\begin{aligned} f_{Y|X}(y|x) &= f_U(r(y)-g(x))|J(y)| \text{ and} \\ f_{Y|X}(y_n|x) &= f_U(r(y_n)-g(x))|J(y_n)|. \end{aligned}$$

By Assumptions D2 and D4, the continuity statement in Proposition 1(i) holds for given $x \in \mathcal{X}$ and $y \in \left(\cup_{l \in L_{s^*}} A_{s^*l}\right) \cap \operatorname{int}(\mathcal{Y}(x))$.

Now suppose $y \in \left(\cap_{k \in K} \bar{A}_{s^*l_k}\right) \cap \operatorname{int}(\mathcal{Y}(x))$, where $K \subset L_{s^*}$ and $\#(K) \geq 2$. Let $O$ be an open set such that $y \in O$ and $O \cap A_{s^*l_k} \neq \varnothing$ for all $k \in K$. Then for any $y' \in O \cap A_{s^*l_k} \cap \operatorname{int}(\mathcal{Y}(x))$, applying the result of last paragraph leads to

$$f_{Y|X}(y'|x) = f_U(r(y') - g(x))|J(y')|.$$

Consider any sequence $\{y'_n\}$ such that $y'_n \in O \cap A_{s^*l_k} \cap \operatorname{int}(\mathcal{Y}(x))$ and $y'_n \to y$, Assumptions D2 and D4 lead to

$$\lim_{O \cap A_{s^*l_k} \cap \operatorname{int}(\mathcal{Y}(x)),\ y'_n \to y'} f_{Y|X}(y'_n|x) = f_U(r(y') - g(x))|J(y')|.$$

Thus the continuity statement in Proposition 1(i) holds for any $y \in \left(\cap_{k \in K} \bar{A}_{s^*l_k}\right) \cap \operatorname{int}(\mathcal{Y}(x))$ and some $K \subset L_{s^*}$ with $\#(K) \geq 2$.

Using similar arguments, we can show that the continuity statement in Proposition 1(i) holds for any $y \in \left(\cap_{s \in S} \bar{A}_s\right) \cap \operatorname{int}(\mathcal{Y}(x))$ and some $S \subset \{l, ..., M\}$ with $\#(S) \geq 2$.

Note that any $y \in \operatorname{int}(\mathcal{Y}(x))$ must belong to one of the three sets: $\cup_{l \in L_{s^*}} A_{s^*l}$, $\cap_{k \in K} \bar{A}_{s^*l_k}$ and $\cap_{s \in S} \bar{A}_s$. Thus the desired result is proven.

(ii) For any $x \in \mathcal{X}$ and $y \in \left(\cup_{l \in L_{s^*}} A_{s^*l})\right) \cap \operatorname{int}(\mathcal{Y}(x))$, we have

$$\lambda_{s^*}(x, r(y) - g(x)) = 1. \tag{A.2}$$

Consider any sequence $\{y_n, x_n\}$ such that $y_n \in \operatorname{int}(\mathcal{Y}(x_n))$ and $(y_n, x_n) \to (y, x)$, we must have

$$\lambda_{s^*}(x_n, r(y_n) - g(x_n)) = 1,$$

for sufficiently large $n$, where $s^*$ is the same as the one in (A.2). Suppose it is not. Then for any $N$, there exists an $n \geq N$ such that $\lambda_{s^*}(x_n, r(y_n) - g(x_n)) = 0$. For $(y_n, x_n)$ arbitrarily close to $(y, x)$, $y_n \in B_\epsilon(y) \subset A_{s^*}$ for arbitrarily small $\epsilon$. By Lemma 2, $f_{Y|X}(y_n|x_n) = 0$, which contradicts with $y_n \in \operatorname{int}(\mathcal{Y}(x_n))$.

Then by Lemma 2, we have

$$\begin{aligned} f_{Y|X}(y|x) &= f_U(r(y)-g(x))|J(y)| \text{ and} \\ f_{Y|X}(y_n|x_n) &= f_U(r(y_n)-g(x_n))|J(y_n)|. \end{aligned}$$

Thus the continuity statement in Proposition 1(ii) follows from Assumptions D2 and D4.

Arguments similar to part (i) applies for $y \in \left(\cap_{k\in K}\bar{A}_{s^*l_k}\right) \cap \text{int}(\mathcal{Y}(x))$, $K \subset L_{s^*}$ with $\#(K) \geq 2$, and for $y \in \left(\cap_{s\in S}\bar{A}_s\right) \cap \text{int}(\mathcal{Y}(x))$, $S \subset \{l, ..., M\}$ with $\#(S) \geq 2$. This completes the proof. **Q.E.D.**

**Proof of Proposition 2.** (i) Suppose $A_{sh}$ and $A_{sh'}$ satisfy Assumption D6 (i). For any $y \in A_{sh}$ and $y' \in A_{sh'}$, Lemma 2 gives

$$\begin{aligned} f_{Y|X}(y|x) &= f_U(r(y)-g(x))|J(y)|\lambda_s(x, r(y)-g(x)) \text{ and} \\ f_{Y|X}(y'|x) &= f_U(r(y')-g(x))|J(y')|\lambda_s(x, r(y')-g(x)). \end{aligned}$$

Let $y_d \in \bar{A}_{sh} \cap \bar{A}_{sh'}$ and $x \in \mathcal{X}_D$. Observe that

$$\begin{aligned} \lim_{y\in A_{sh}, y\to y_d} f_{Y|X}(y|x) &= f_U(r(y_d)-g(x))|J(y_d)| \\ &\quad \lim_{y\in A_{sh}, y\to y_d} \lambda_s(x, r(y)-g(x)), \\ \lim_{y\in A_{sh'}, y'\to y_d} f_{Y|X}(y'|x) &= f_U(r(y_d)-g(x))|J(y_d)| \\ &\quad \lim_{y\in A_{sh'}, y'\to y_d} \lambda_s(x, r(y')-g(x)). \end{aligned} \tag{A.3}$$

Since

$$\lim_{y\in A_{sh}, y\to y_d} \lambda_s(x, r(y)-g(x)) \neq \lim_{y\in A_{sh'}, y'\to y_d} \lambda_s(x, r(y')-g(x)),$$

and both are non-zero by Assumption D6(ii). Together with $f_U(r(y_d) - g(x)) > \eta > 0$ by the fact that $y_d \in \bar{\mathcal{Y}}$ and thus $r(y_d) - g(x) \in \mathcal{U}$, and $|J(y_d')| > \varsigma$ by Assumption D5(i), we have $\lim_{y\in A_{sh}, y\to y_d} f_{Y|X}(y|x) \neq 0$ and $\lim_{y\in A_{sh'}, y'\to y_d} f_{Y|X}(y'|x) \neq 0$ as well as

$$\begin{aligned} & \left| \lim_{y\in A_{sh}, y\to y_d} f_{Y|X}(y|x) - \lim_{y\in A_{sh'}, y'\to y_d} f(y'|x) \right| \\ = \; & f_U(r(y_d)-g(x))|J(y_d)| \\ \times \; & \left| \lim_{y\in A_{sh}, y\to y_d} \lambda_s(x, r(y)-g(x)) - \lim_{y\in A_{sh'}, y'\to y_d} \lambda_s(x, r(y')-g(x)) \right| > 0. \end{aligned}$$

Therefore, $f_{Y|X}(y|x)$ has a jump at $y_d$ for any $x \in \mathcal{X}_D$.

(ii) immediately follows from (i). **Q.E.D.**

To prove Corollary 1, let us first state Lemma 3 as follows.

**Lemma 3.** *Suppose that Assumptions D5(ii) and (iii) hold and* $M > 1$. *Then*

*(i) For any* $m \in \{1, ..., M\}$, *there exists a* $k \in \{1, ..., M\}$ *and* $k \neq m$, *such that* $B_m \cap B_k \neq \varnothing$.

*(ii) For any* $m$, *there exists an* $k \neq m$ *such that* $r(A_m) \cap r(A_k) \neq \varnothing$.

**Proof of Lemma 3.**

(i). It suffices to show that for every $m \in \{1, ..., M\}$, there exists some $v \in B_m$ such that $r(y) = v$ holds for more than one $y \in \bar{\mathcal{Y}}$. Suppose it is not true. Then we have a $B_m$ such that $r(y) = v$ has a unique solution in $y$ for all $v \in B_m$. Consider an $l \in \{j \neq m : \bar{B}_m \cap \bar{B}_j \neq \varnothing\}$ and let $v_0$ be an arbitrary point $v_0 \in \bar{B}_m \cap \bar{B}_l$. By assumption, $B_m \cap B_l = \varnothing$ for all $l \neq m$. Thus $v_0 \in \bar{B}_m \backslash B_m$. For any such $l$, we have two cases:

Case 1. $\lim_{v \in B_m, v \to v_0} q_m(v) - \lim_{v \in B_l, v \to v_0} q_l(v) = 0$ for all $v_0 \in \bar{B}_m \cap \bar{B}_l$.

In this case, we can combine $B_m$ and $B_l$ to reduce the number $M$, this contradicts with the Definition 1.

In particular, let $B'_m = \text{int}\big(B_m \cup B_l \cup \big(\bar{B}_m \cap \bar{B}_l\big)\big)$ and

$$q'_m(v) = \begin{cases} q_m(v), & v \in B_m, \\ q_l(v), & v \in B_l, \\ \lim_{v \in B_m, v \to v_0} q_m(v), & v_0 \in \bar{B}_m \cap \bar{B}_l. \end{cases}$$

By Assumption D5(ii), we know that $\triangledown r(y_0)$ is invertible, where $y_0 = \lim_{v \in B_m, v \to v_0} q_m(v) = \lim_{v \in B_l, v \to v_0} q_l(v)$. Then we can apply the Implicit Function Theorem to obtain that there are neighbourhoods $A$ of $y_0$ and $B$ of $v_0$, and a unique function $\varsigma : B \to A$ such that $\varsigma(v_0) = y_0$ and for each $v \in B$, $\varsigma(v)$ is the unique solution to $r(y) = v$ lying in $A$, $r(\varsigma(v)) = v$ for $v \in B$, and $\varsigma(v)$ is twice continuously differentiable at $x_0$. Therefore, $q'_m(v)$ must coincide with $\varsigma(v)$ for $v \in B$. Since $v_0$ is an arbitrary point in $\bar{B}_m \cap \bar{B}_l$, $q'_m(v)$ is twice continuously differentiable on $B'_m$.

Case 2. $|\lim_{v \in B_m, v \to v_0} q_m(v) - \lim_{v \in B_l, v \to x_0} q_l(v)| = \delta > 0$, for some $v_0 \in \bar{B}_m \cap \bar{B}_l$.

In this case, by Assumption D2, $r(y)$ is continuous on $\bar{\mathcal{Y}}$. Hence

$$
\begin{aligned}
r(y_m) &= r(y_l) = v_0, \\
\text{where } y_m &= \lim_{v \to v_0, v \in B_m} q_m(x), \; y_l = \lim_{v \to v_0, v \in B_l} q_l(x).
\end{aligned}
$$

There must be an $l$ satisfying the following: there is no $l'$ such that every component of $y_{l'} - y_m$ has the same sign as $y_l - y_m$ and $|y_{l'} - y_m| < |y_l - y_m|$. We focus on such an $l$. By Assumption D5(iii), derivative $\bigtriangledown r(y_l)$ is invertible. By applying the Implicit Function Theorem, for any neighbourhood $A$ of $y_l$, there is a neighbourhood $B$ of $v_0$ and a function $\xi : B \to A$ such that $\xi(v_0) = y_l$, $r(\xi(v)) = v$ for $v \in B$, and $\xi(x)$ is continuous at $v_0$. Therefore, for any $\epsilon$, there exists an $\eta$ such that $|v - v_0| < \eta \Longrightarrow |\xi(v) - y_l| < \epsilon$. Then for any $\epsilon$, all $v' \in B_m \cap B(v_0, \eta) \cap B$ satisfy $r(\xi(v')) = v'$ and $|\xi(v') - y_l| < \epsilon$. Thus $|\xi(v') - y_m| > \delta - \epsilon$ and $r(\xi(v')) = v'$ for all $v' \in B_m \cap B(v_0, \eta) \cap B$. On the other hand, by continuity of $q_m$, for any $\epsilon'$, there exists a $\eta'$ such that $|v - v_0| < \eta' \Longrightarrow |q_m(v) - y_m| < \epsilon'$. Therefore, for any $\epsilon'$, all $v'' \in B_m \cap B(v_0, \eta')$ satisfy $r(q_m(v'')) = v''$ and $|q_m(v'') - y_m| < \epsilon'$. By choosing $\epsilon$ and $\epsilon'$ sufficiently small, we have $\xi(v) \neq q_m(v)$ and $r(\xi(v)) = r(q_m(v)) = v$, for all $v \in B_m \cap B(v_0, \eta \wedge \eta') \cap B$. This contradicts with the assumption that $r(y) = v$ has a unique solution in $y$ for all $v \in B_m$.

(ii) Immediately follows from (i) by letting $A_m = q_m(B_m)$. **Q.E.D.**

**Proof of Corollary 1.**

By Assumption D7(i), there exists an $s \in \{1, 2, ..., M\}$ such that $\Pr(g(X) + U \in r(A_s) \cap S) > 0$.

Let $A_{s1} = \{y \in A_s : r(y) \in S\}$. We have $(\mathcal{M}(r(A_{s1}))) = \{s\}$. Also By Lemma 3 (ii), there exists some $m \neq s$ such that $r(A_s) \cap r(A_m) \neq \varnothing$. Let $A_{s2} = \{y \in A_s : r(y) \in r(A_m)\}$. We have $\{s, m\} \subset \mathcal{M}(r(A_{s2}))$. Thus $\mathcal{M}(r(A_{s1})) \neq \mathcal{M}(r(A_{s2}))$. Assuming $\bar{A}_{s1} \cap \bar{A}_{s2} \neq \varnothing$ does not lose any generality. **Q.E.D.**

**Proof of Corollary 2.**

Under Assumptions D8(i) and D7(ii), there exists some $k \in L_m$ and $C_0 \subset A_{mk}$ such that for all $x \in \mathcal{X}_D$,

$$0 < \phi_m(\mathcal{M}(r(C_0)), x) = \phi_m(\mathcal{M}(r(A_{mk})), x) < 1.$$

By Assumption D7(i), there exists some $j \in L_m$ such that $\#(\mathcal{M}(r(A_{mj})))$ $= 1$. Thus $\phi_m(\mathcal{M}(r(A_{mj})), x) = 1$. There are two cases:

Case 1. $\bar{A}_{mj} \cap \bar{A}_{mk} \neq \varnothing$, the rest of proof follows that of Proposition 2, and $\mathcal{X}_{D_1} = \mathcal{X}_D$.

Case 2. $\bar{A}_{mj} \cap \bar{A}_{mk} = \varnothing$, that means for all $h$ such that $\bar{A}_{mh} \cap \bar{A}_{mk} \neq \varnothing$, $\#(\mathcal{M}(r(A_{mh}))) > 1$:

Case 2.1. $\phi_m(\mathcal{M}(r(A_{mh})), x) \neq \phi_m(\mathcal{M}(r(A_{mk})), x)$ for some $h$ such that $\bar{A}_{mh} \cap \bar{A}_{mk} \neq \varnothing$, and for a subset $\mathcal{X}_{D_1} \subset \mathcal{X}_D$. Then the rest of proof follows that of Proposition 2.

Case 2.2. $\phi_m(\mathcal{M}(r(A_{mh})), x) = \phi_m(\mathcal{M}(r(A_{mk})), x)$ for all $h$ such that $\bar{A}_{mh} \cap \bar{A}_{mk} \neq \varnothing$, and for almost all $x$. Then we must have $0 < \phi_m(\mathcal{M}(r(A_{mh})), x) < 1$, for all $h$ such that $\bar{A}_{mh} \cap \bar{A}_{mk} \neq \varnothing$. Consequently, we can replace $k$ with an $h$ such that $\bar{A}_{mh} \cap \bar{A}_{mk} \neq \varnothing$, and repeat the procedure from the beginning of the proof. Such a procedure continues until an end in either Case 1 or Case 2.1, because $\#(L_m)$ is finite. **Q.E.D.**

**Proof of Corollary 3.**

By Assumption D7(i), without loss of generality, assume $A_1 \cap S \neq \varnothing$ and $A_{11} = \{y \in A_1 : \mathcal{M}(r(y)) = \{1\}\}$. Note that we also have $A_1 = \cup_{l \in L_1} \bar{A}_{1l}$.

If the model is in (C3), by Corollary 1, there exists some $A_{12}$ such that $\mathcal{M}(r(A_{12})) \neq \mathcal{M}(r(A_{11}))$. Since $\mathcal{M}(r(A_{12}))$ includes equilibrium 1 by definition, we must have $\#(\mathcal{M}(r(A_{12}))) > 1$. Suppose $\bar{A}_{11} \cap \bar{A}_{12} \neq \varnothing$. (This does not reduce generality since Category (C3) and Assumption D7(i) imply that there exist $A_{1j}$ and $A_{1k}$ satisfying $\bar{A}_{1j} \cap \bar{A}_{1k} \neq \varnothing$ and $\mathcal{M}(r(A_{1j})) = 1$, $\mathcal{M}(r(A_{1k})) > 1$). Given an arbitrary $x$, for any $y \in \bar{A}_{11} \cap \bar{A}_{12}$, by Lemma 2 we have

$$\begin{aligned}
\lim_{y' \to y, y' \in A_{11}} f_{Y|X}(y'|x) &= f_U(r(y) - g(x))|J(y)|\phi_1(\mathcal{M}(r(A_{11})), x) \\
&= f_U(r(y) - g(x))|J(y)| \qquad \text{and} \\
\lim_{y' \to y, y' \in A_{12}} f_{Y|X}(y'|x) &= f_U(r(y) - g(x))|J(y)|\phi_1(\mathcal{M}(r(A_{12})), x),
\end{aligned}$$

where $\phi_1(\mathcal{M}(r(A_{11})), x) = 1$ since $\mathcal{M}(r(A_{11})) = \{1\}$. By Assumption D9, we have the following cases:

Case 1. There is an $\mathcal{X}_{D_2} \subset \mathcal{X}$ with $\Pr(X \in \mathcal{X}_{D_2}) > 0$ such that $0 < \phi_1(\mathcal{M}(r(A_{12})), x) < 1$ for all $x \in \mathcal{X}_{D_2}$.

By Assumption D5(i) we have

$$\begin{aligned}
&\left| \lim_{y' \to y, y' \in A_{11}} f_{Y|X}(y'|x) - \lim_{y' \to y, y' \in A_{21}} f_Y(y'|x) \right| \\
= \;& f_U(r(y) - g(x))|J(y)|(1 - \phi_1(\mathcal{M}(r(A_{12})), x))
\end{aligned}$$

$$\geq \quad \eta\delta\varsigma > 0.$$

That is, $f_{Y|X}(y|x)$ has a jump at any $y \in \bar{A}_{11} \cap \bar{A}_{12}$ for some $x \in \mathcal{X}_{D_2}$. Furthermore, D5(i) and D9(ii) makes sure that $\lim_{y' \to y, y' \in A_{11}} f_{Y|X}(y'|x) \neq 0$ and $\lim_{y' \to y, y' \in A_{12}} f_{Y|X}(y'|x) \neq 0$. The desired result is obtained.

Case 2.1. If $\phi_1(\mathcal{M}(r(A_{12})), x) = 1$ for almost all $x$.

Consider another $A_{1j}$ such that $\bar{A}_{11} \cap \bar{A}_{1j} \neq \varnothing$, by the same argument as in Case 1, $f_{Y|X}(y|x)$ has a jump at $y \in \bar{A}_{11} \cap \bar{A}_{1j}$, if there exists an $\mathcal{X}_{D_2}$ such that $0 < \phi_1(\mathcal{M}(r(A_{1j})), x) < 1$ for $x \in \mathcal{X}_{D_2}$. Repeating this step yields the desired result if there exists some $l \in L_1$ such that $0 < \phi_1(\mathcal{M}(r(A_{1l})), x) < 1$ for $x \in \mathcal{X}_{D_2}$.

Case 2.2. If $\phi_1(\mathcal{M}(r(A_{1l})), x) = 1$ for almost all $x$ and all $l \in L_1$.

Consider some other $m \neq 1$ with $r(A_m) \cap S \neq \varnothing$. By repeating the steps in Case 1 and Case 2.1, the desired result is obtained if there exists some $m$ such that $0 < \phi_m(\mathcal{M}(r(A_{ml})), x) < 1$, and for some $l \in L_m$ and for all $x \in \mathcal{X}_{D_2}$ .

Case 2.3. If $\phi_m(\mathcal{M}(r(A_{ml})), x) = 1$ for almost all $x$, all $l \in L_m$ and all $m$ with $r(A_m) \cap S \neq \varnothing$.

Let $T = \{t : r(A_t) \cap S \neq \varnothing\}$. Case 2.3 means that for almost all $x$, there exists a unique $t \in T$ ($t$ may depend on $x$) such that $\phi_t(\mathcal{M}(r(A_{tl})), x) = 1$, all $l \in L_t$ and $\phi_v(\mathcal{M}(r(A_{vk})), x) = 0$ for any $v \notin T$ and any $k \in L_v$. By Assumption D9(i), for any $(x, u)$, $g(x) + u \in r(A_t)$ for some $t \in T$. Thus by Assumotion D7(ii), the equilibrium selection rule is degenerate (always choose equilibrium $t$ with probability 1) and definition of (C3) is violated. Hence Case 2.3 cannot happen when there are multiple equilibria. **Q.E.D.**

**Proof of Proposition 3.**

(i) By Assumption D10, given any $m$,

$$\lambda_m(x, u) = \psi_m(g(x) + u) = \psi_m(r(y)) = 1 \text{ or } 0.$$

For any $y \in \cup_{m=1}^{M} \cup_{l \in L_m} A_{ml}$, by Lemma 2 and Proposition 1(i), we have for all $x \in \mathcal{X}$,

$$f_{Y|X}(y|x) = f_U\left(r(y) - g(x)\right)|J(y)|\psi_m(r(y)),$$

where $m$ and $l$ satisfy $y \in A_{ml}$. Observe that

$$\begin{aligned} f_Y(y) &= \int_{\mathcal{X}} f_U\left(r(y) - g(x)\right)|J(y)|\psi_m(r(y))dF_X(x) \\ &= \psi_m(r(y))|J(y)| \int_{\mathcal{X}} f_U\left(r(y) - g(x)\right)dF_X(x) \end{aligned}$$

$$= \begin{cases} |J(y)| \int_{\mathcal{X}} f_U(r(y) - g(x))\, dF_X(x) & \text{if } \psi_m(r(y)) = 1 \\ 0 & \text{if } \psi_m(r(y)) = 0 \end{cases},$$

which is continuous as long as $y \in int(\mathcal{Y}(x))$ by Assumptions D2 and D4. Moreover, the density is

$$|J(y)| \int_{\mathcal{X}} f_U(r(y) - g(x))\, dF_X(x).$$

Since the choice of $m$ is arbitrary and $\bar{\mathcal{Y}} = y \in \cup_{m=1}^{M} \cup_{l \in L_m} \bar{A}_{ml}$, the desired result is proven.

(ii) Given the assumptions, the conclusion of Proposition 2(i) holds. From By Assumption D5 and D7, there is a jump $y_d \in \bar{A}_{mh} \cap \bar{A}_{mk}$ for $x \in \mathcal{X}_D$, where $\#(\mathcal{M}(r(A_{mh}))) = 1$ and $\#(\mathcal{M}(r(A_{mh}))) > 1$. Thus the jump size

$$\lim_{y \in A_{mh}, y \to y_d} f_{Y|X}(y|x) - \lim_{y \in A_{mk}, y \to y_d} f_{Y|X}(y|x)$$

at an arbitrary $x \in \mathcal{X}_D$ is

$$f_U(r(y_d) - g(x))\, |J(y_d)| \left[1 - \lim_{y \in A_{mk}, y \to y_d} \phi_m(\mathcal{M}(r(A_{mk})), x)\right] = \delta > 0.$$

Therefore, the difference in the unconditional density can be written as

$$\begin{aligned} & \lim_{y \in A_{mh}, y \to y_d} f_Y(y) - \lim_{y \in A_{mk}, y \to y_d} f_Y(y) \\ = & \int_{\mathcal{X}_D} \left[\lim_{y \in A_{mh}, y \to y_d} f_{Y|X}(y|x) - \lim_{y \in A_{mk}, y \to y_d} f_{Y|X}(y|x)\right] f(x)dx \\ > & \ \delta \Pr(X \in \mathcal{X}_D) > 0. \end{aligned}$$

Also, since $\lim_{y \in A_{mh}, y \to y_d} f_{Y|X}(y|x) \neq 0$ for $x \in \mathcal{X}_D$, we have

$$\lim_{y \in A_{mh}, y \to y_d} f_Y(y) = \int_{\mathcal{X}} f_{Y|X}(y|x) \geq \int_{\mathcal{X}_D} f_{Y|X}(y|x) \neq 0.$$

The same is for $\lim_{y \in A_{mk}, y \to y_d} f_Y(y)$.**Q.E.D.**

**Proof of Proposition 4.**

Let $\mathcal{U}_D$ denote the collection of jump locations of $f_U(u)$. If the model has a unique equilibrium, by Assumption D10, given any $m$, $\lambda_m(x,u) = \psi_m(g(x) + u) = \psi_m(r(y)) = 1$ or 0.

For an arbitrary $y \in \cup_{m=1}^{M} \cup_{l \in L_m} A_{ml}$, by Lemma 2,

$$f_{Y|X}(y|x) = f_U(r(y) - g(x))\, |J(y)| \psi_m(r(y)),$$

for all $x \in \mathcal{X} \backslash \mathcal{X}_{JL}(y)$, where $\mathcal{X}_{JL}(y) = \{x : r(y) - g(x) \in \mathcal{U}_D\}$, $m$ and $l$ satisfy $y \in A_{ml}$. Since $\Pr(U \in \mathcal{U}) = 0$ and Assumption D11 holds, $\Pr(X \in \mathcal{X}_{JL}(y)) = 0$, for all $y$. Hence for any arbitrary $y \in \cup_{m=1}^{M} \cup_{l \in L_m} A_{ml}$,

$$\begin{aligned}
f_Y(y) &= \int_{\mathcal{X} \backslash \mathcal{X}_{JL}(y)} f_U\left(r(y) - g(x)\right) |J(y)| \psi_m(r(y)) dF_X(x) \\
&\quad + \int_{\mathcal{X}_{JL}(y)} f_{Y|X}(y|x) dF_X(x) \\
&= \psi_m(r(y)) |J(y)| \int_{\mathcal{X} \backslash \mathcal{X}_{JL}(y)} f_U\left(r(y) - g(x)\right) dF_X(x) \\
&= \begin{cases} |J(y)| \int_{\mathcal{X} \backslash \mathcal{X}_{JL}(y)} f_U\left(r(y) - g(x)\right) dF_X(x) & \text{if } \psi_m(r(y)) = 1 \\ 0 & \text{if } \psi_m(r(y)) = 0 \end{cases},
\end{aligned}$$

which is continuous at as long as $y \in int(\mathcal{Y}(x))$ by Assumptions D2 and D4. Moreover, the density is

$$|J(y)| \int_{\mathcal{X}} f_U\left(r(y) - g(x)\right) dF_X(x).$$

Since the choice of $m$ is arbitrary and $\bar{\mathcal{Y}} = y \in \cup_{m=1}^{M} \cup_{l \in L_m} \bar{A}_{ml}$, the desired result follows. **Q.E.D.**

## A.2 Conditions in Section 2.4

This section spells out five conditions (SL1 to SL5) sufficient for Proposition 5 and Proposition 6 in Section 2.4. Note that Conditions SL1 to SL4 are adapted from Chernozhukov et al. (2013).

To state those conditions, let us first define the population version of the aforementioned maximizing set $\hat{\mathcal{Y}}_{\kappa_2}$. Let

$$\mathcal{Y}_{\kappa_2} = \{y \in \mathcal{Y} : P_j^-(y)/P_j(y) > \kappa_2, P_j^+(y)/P_j(y) > \kappa_2, \text{ all } j \in \mathcal{J}\},$$

where for $j \in \mathcal{J}$,

$$\begin{aligned}
P_j^-(y) &= \Pr\left(y_j - w_j \le Y_{i,j} < y_j | Y_{i,-j} = y_{-j}\right) f_{Y_{-j}}\left(y_{-j}\right), \\
P_j^+(y) &= \Pr\left(y_j \le Y_{i,j} < y_j + w_j | Y_{i,-j} = y_{-j}\right) f_{Y_{-j}}\left(y_{-j}\right), \\
P_j(y) &= P_j^+(y) + P_j^-(y), \ w_j = (\sup_y \mathcal{Y}_j - \inf_y \mathcal{Y}_j)/H.
\end{aligned}$$

Conditions SL1 and SL2 are imposed on the kernel function. For $j \in \mathcal{J}$, recall that $\Delta\mathcal{K}_{n,j}$ be a class of measurable functions,

$$\Delta\mathcal{K}_{n,j} = \left\{ \frac{\sqrt{p_n h_n^{J-1}}\varphi_n^{(j)}(\cdot, y)}{\sigma_{j,n}(y)} : y \in \mathcal{Y}_{\kappa_2} \right\}, \tag{A.4}$$

where

$$\varphi_n^{(j)}(t,y) = \frac{1}{p_n h_n^{J-1}} \left[ \; K_j^+ \left( \tfrac{t_j - y_j}{p_n}, \tfrac{t_{-j} - y_{-j}}{h_n} \right) - K_j^- \left( \tfrac{t_j - y_j}{p_n}, \tfrac{t_{-j} - y_{-j}}{h_n} \right) \; \right]. \tag{A.5}$$

Recall that the covering number $N\left(\Delta\mathcal{K}_{n,j}, L_2(Q), b_{n,j}\varepsilon\right)$ is defined as the minimum number of closed $L_2$ balls of radius $b_{n,j}\varepsilon$ to cover $\Delta\mathcal{K}_{n,j}$.

**Condition SL1.** (i) $\Delta\mathcal{K}_{n,j}$ is uniformly bounded from above by a number $b_{n,j}$, and the covering number of $\Delta\mathcal{K}_{n,j}$ satisfies

$$\sup_Q N\left(\Delta\mathcal{K}_{n,j}, L_2(Q), b_{n,j}\varepsilon\right) \leq \left(a_j/\varepsilon\right)^{v_j}, \text{ for all } 0 < \varepsilon < 1, \tag{A.6}$$

for some $a_j > e$ and $v_j \geq 1$, where the supermum is taken over all finitely discrete probability measures $Q$.

(ii) For any $g \in \Delta\mathcal{K}_{n,j}$, there exists a constant $\sigma_{n,j}$ such that $\mathbf{E}\left[g(Y_1)^2\right] \leq \sigma_{n,j}^2 \leq b_{n,j}^2$.

**Condition SL2.** There exist positive constants $c_2$ and $C_2$ such that for all $j \in \mathcal{J}$,

$$\begin{aligned} c_2 &\leq \inf_{y \in \mathcal{Y}_{\kappa_2}} \sigma_{j,n}(y) \leq \sup_{y \in \mathcal{Y}_{\kappa_2}} \sigma_{j,n}(y) \leq C_2, \\ c_2 &\leq \sigma_{n,j}^2 \leq C_2. \end{aligned}$$

**Condition SL3.** Suppose that $H_0$ is true, or $H_1$ is true and for any $y \in \mathcal{Y}_{\kappa_2}$, there exists $\varepsilon$ such that $f(\cdot)$ is twice continuously differentiable over a neighbourhood of $(y_j + \varepsilon, y_{-j})$ and $(y_j - \varepsilon, y_{-j})$ for all $j \in \mathcal{J}$. Then there are integer $n_0$ and positive constants $c_3, C_3, c_4$ and $C_4$ such that for all $n \geq n_0$, there is some $t$ satisfying $c_3 \leq t \leq C_3$ and

$$\sup_{y \in \mathcal{Y}_{\kappa_2}} \left( \mathbf{E}\left[\Delta_n^{(j)}(y)\right] - \Delta^{(j)}(y) \right) \leq C_4 h_n^t,$$

$$\sup_{y\in\mathcal{Y}_{\kappa_2}} \left(\mathbf{E}\left[\Delta_n^{(j)}(y)\right] - \Delta^{(j)}(y)\right) \quad \leq \quad C_4 p_n^t.$$

**Condition SL4.** Let $K_n = Av(\log n \vee \log(ab_n/\sigma_n))$, where $b_n = \max_{j\in\mathcal{J}} b_{n,j}$, $\sigma_n = \max_{j\in\mathcal{J}} \sigma_{n,j}$, $v = \max_{j\in\mathcal{J}} v_j$ and $a = J\max_{j\in\mathcal{J}} a_j$, $A$ is a constant. Then there are constants $C_5$ and $c_5$ such that $b_n K_n^4/n \leq C_5 n^{-c_5}$, and a constant $\bar{M} > 0$ such that

$$\sqrt{np_n h_n^{J-1}} p_n^t \quad \leq \quad \bar{M} c_{1,n}^S(\gamma_n) \sup_{j\in\mathcal{J},y\in\mathcal{Y}_{\kappa_2}} \sigma_{j,n}(y),$$

$$\sqrt{np_n h_n^{J-1}} h_n^t \quad \leq \quad \bar{M} c_{1,n}^S(\gamma_n) \sup_{j\in\mathcal{J},y\in\mathcal{Y}_{\kappa_2}} \sigma_{j,n}(y).$$

**Condition SL5.** Let bandwidths be $p_n = p_0 \times n^{-\gamma_p}$ and $h_n = h_0 \times n^{-\gamma_h}$. Assume that the bandwidths satisfy $\gamma_p + (J-1)\gamma_h < 1$, $\gamma_p + (J+1)\gamma_h > 1$ and $3\gamma_p + (J-1)\gamma_h > 1$.

The constant $t$ in Condition SL3 differs under $H_0$ and $H_1$. Conditions SL4 and SL5 place restrictions on the bandwidths $p_n$ and $h_n$. Note that under Condition SL5, $\sqrt{np_n h_n^{J-1}} p_n^t$ and $\sqrt{np_n h_n^{J-1}} h_n^t$ on the left hand side of Condition SL4 converge to zero. The requirement of $C$ in computing the critical value is $C \geq \bar{M}C_4/((1-\epsilon_{3n})c_2)$ (see Appendix A.3). Then under Condition SL5, the constant $C$ can be chosen arbitrarily close to zero. In practice, we can set $C = 0$.

## A.3 Proofs for Section 2.4

The organization of this Appendix A.3 is as follows. We first introduce some notations. Then we spell out four high level conditions in the spirit of Chernozhukov et al. (2013), and verify them under Conditions SL1-SL5 in Appendix A.2. Next we state and prove three lemmas. Lastly, we prove Propositions 5 and 6 using high level conditions as well as the lemmas.

### Notations

Let $c_{1,n}^S(\alpha)$ be the $(1-\alpha)$th quantile of the distribution of $\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{\kappa_2}} \left|\hat{\mathbb{G}}_{n,j}(y)\right|$ given $\{Y_i\}_{i=1}^n$. Also let $c_{1,n}^S = c_{1,n}^S(\alpha) + c'_{1,n}$, where

$c'_{1,n} = Cc_{1,n}(\gamma_n)$, and $c_{1,n}(\gamma_n)$ is the $(1-\gamma_n)$th quantile of the distribution of $\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{\kappa_2}}\left|\hat{\mathbb{G}}_{n,j}(y)\right|$.

Let $Z_{j,n}$ be an empirical process

$$\begin{aligned} Z_{j,n} &= \left\{Z_{j,n}(y) : j\in\mathcal{J}, y\in\mathcal{Y}_{\kappa_2}\right\}, \\ \text{where } Z_{j,n}(y) &= \frac{\sqrt{np_nh_n^{J-1}}\left(\Delta_n^{(j)} - \mathbf{E}\left[\Delta_n^{(j)}(y)\right]\right)}{\sigma_{j,n}(y)}. \end{aligned}$$

$Z_{j,n}$ can also be written as

$$Z_n(g) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(g(Y_i) - \mathbf{E}\left[g(Y_i)\right]\right),\ g\in\Delta\mathcal{K}_n,$$

where $\Delta\mathcal{K}_n \equiv \cup_{j=1}^{J}\Delta\mathcal{K}_{n,j}$ defined in (A.4).

Let $G = \{G(g) : g\in\Delta\mathcal{K}_n\}$ be a zero mean Gaussian process with the same covariance function as that of $Z_n = \{Z_n(g) : g\in\Delta\mathcal{K}_n\}$. That is,

$$\begin{aligned} & \mathbf{E}\left[G(g_1)G(g_2)\right] \\ = & \mathbf{E}\left[Z_n(g_1)Z_n(g_2)\right] - \mathbf{E}\left[Z_n(g_1)\right]\mathbf{E}\left[Z_n(g_2)\right] \\ = & p_nh_n^{J-1}\left[\mathbf{E}\left[g_1(Y_i)g_2(Y_i)\right] - \mathbf{E}\left[g_1(Y_i)\right]\mathbf{E}\left[g_2(Y_i)\right]\right]. \end{aligned} \tag{A.7}$$

for any $g_1, g_2\in\Delta\mathcal{K}_n$.

**High level conditions**

We state following four high level conditions analogous to Chernozhukov et al. (2013). These conditions will be used in the proof of Propositions 5 and 6.

**Condition SH1.** One can construct a sequence of random variables $W_n^0$ satisfying the following conditions.

(i) $W_n^0 =_d \sup_{g\in\Delta\mathcal{K}_n}|G(g)|$, where $G(g)$ is a centred Gaussian process with $\mathbf{E}\left[G^2(g)\right] = 1$ for any $g\in\Delta\mathcal{K}_n$, and $\mathbf{E}\left[\sup_{g\in\Delta\mathcal{K}_n}|G(g)|\right] \le C_1\sqrt{\log n}$.

(ii) For some positive sequences $\epsilon_{1n}$ and $\delta_{1n}$ bounded by $C_1n^{-c_1}$ (for some constants $C_1, c_1$). We have

$$\Pr\left(\left|\sup_{g\in\Delta\mathcal{K}_n}|Z_n(g)| - W_n^0\right| > \epsilon_{1n}\right) \le \delta_{1n}. \tag{A.8}$$

**Condition SH2.** Let $c_n$ be the $(1-\alpha)$th quantile of the distribution of $\sup_{g \in \Delta \mathcal{K}_n} |G(g)|$, for some positive sequences $\tau_n, \epsilon_{2n}$ and $\delta_{2n}$ bounded by $C_1 n^{-c_1}$, we have

$$\begin{aligned}
\Pr\left(c^S_{1,n}(\alpha) < c_n(\alpha+\tau_n) - \epsilon_{2n}\right) &\leq \delta_{2n}, \\
\Pr\left(c^S_{1,n}(\alpha) > c_n(\alpha-\tau_n) + \epsilon_{2n}\right) &\leq \delta_{2n}.
\end{aligned}$$

**Condition SH3.** For some positive sequences $\epsilon_{3n}$ and $\delta_{3n}$ bounded by $C_1 n^{-c_1}$, we have

$$\Pr\left(\sup_{j \in \mathcal{J}, y \in \mathcal{Y}_{\kappa_2}} \left|\frac{\hat{\sigma}_{j,n}(y)}{\sigma_{j,n}(y)} - 1\right| > \epsilon_{3n}\right) \leq \delta_{3n}.$$

For future reference, we state Condition A1 as follows,

**Condition A1.** The alternative hypothesis $H_1$ is true, and for any $y \in \mathcal{Y}_{\kappa_2}$, there is an $\varepsilon$ such that the density function $f(\cdot)$ is twice continuously differentiable over a neighbourhood of $(y_j+\varepsilon, y_{-j})$ and $(y_j-\varepsilon, y_{-j})$ for every $j \in \mathcal{J}$.

**Condition SH4.** Let

$$B_n = \sup_{j \in \mathcal{J}, y \in \mathcal{Y}_{\kappa_2}} \frac{\sqrt{np_n h_n^{J-1}}\mathbf{E}\left[\Delta_n^{(j)}(y) - \Delta^{(j)}(y)\right]}{\hat{\sigma}_{j,n}(y)}.$$

If $H_0$ is true, or Condition A1 holds, then for some positive sequence $\delta_{4n}$ bounded by $C_1 n^{-c_1}$, we have

$$\Pr\left(B_n > c'_{1,n}\right) \leq \delta_{4n}.$$

where $c'_{1,n} = C c^S_{1,n}(\gamma_n)$.

Note that under $H_0$, $\Delta^{(j)}(y)$ is zero for all $j$.

### Verify Conditions SH1-SH4 using Conditions SL1-SL5

Proposition A1 below states that Conditions SH1-SH4 hold under Conditions SL1-SL5.

**Proposition A1**. *Suppose that Conditions SL1-SL4 hold. Then Conditions SH1-SH4 hold.*

**Proof of Proposition A1**

**(i) Verify SH1**: By Condition SL1, $\Delta\mathcal{K}_{n,j}$ is a class of measurable functions uniformly bounded by $b_{n,j}$ and its covering number satisfies

$$\sup_Q N\left(\Delta\mathcal{K}_{n,j}, L_2(Q), b_{n,j}\varepsilon\right) \leq \left(a_j/\varepsilon\right)^{v_j}, \text{ for all } 0 < \varepsilon < 1, \tag{A.9}$$

for some $v_j \geq 1$ and $a_j \geq e$. Then $\Delta\mathcal{K}_n \equiv \cup_{j=1}^J \Delta\mathcal{K}_{n,j}$ is a class of measurable functions uniformly bounded by $b_n = \max_{j\in\mathcal{J}} b_{n,j}$ and its covering number satisfies

$$\sup_Q N\left(\Delta\mathcal{K}_{n,}, L_2(Q), b_n\varepsilon\right) \leq \left(a/\varepsilon\right)^{v}, \ 0 < \varepsilon < 1, \tag{A.10}$$

for some number $v \geq 1, a \geq e$. This is because

$$\begin{aligned}
\sup_Q N\left(\Delta\mathcal{K}_{n,}, L_2(Q), b_n\varepsilon\right) &\leq \sum_{j=1}^J \sup_Q N\left(\Delta\mathcal{K}_{n,j}, L_2(Q), b_n\varepsilon\right) \\
&\leq \sum_{j=1}^J \sup_Q N\left(\Delta\mathcal{K}_{n,j}, L_2(Q), b_{n,j}\varepsilon\right) \\
&\leq J\max_j \left(a_j/\varepsilon\right)^{v_j}\ , 0 < \varepsilon < 1.
\end{aligned}$$

Inequality (A.10) holds by letting $v = \max_{j\in\mathcal{J}} v_j$ and $a = J\max_{j\in\mathcal{J}} a_j$. In the following, We will need Theorem 3.1 of Chernozhukov et al.(2013).

*[Restatement of Theorem 3.1 in Chernozhukov et al.(2013)]*

*Suppose $\mathcal{G}$ is a class of functions uniformly bounded by a constant $b$ such that there exist constants $a \geq e$ and $v > 1$ with $\sup_Q N(\mathcal{G}, L_2(Q), b\epsilon) \leq (a/\epsilon)^v$ for all $0 < \epsilon \leq 1$ . Let $\sigma^2$ be a constant such that $\sup_{g\in\mathcal{G}} \mathbf{Var}(g) \leq \sigma^2 \leq b^2$. Denote $K_n = Av(\log n \vee \log(ab/\sigma))$. Define an empirical process*

$$\mathbb{G}_n(g) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(g(X_i) - \mathbf{E}\left[g(X_i)\right]\right), g \in \mathcal{G}.$$

*Let $W_n = \|\mathbb{G}_n\|_{\mathcal{G}} = \sup_{g\in\mathcal{G}} |\mathbb{G}_n(g)|$, and $B = \{B(g) : g \in \mathcal{G}\}$ be a centred tight Gaussian process with a covariance function*

$$\mathbf{E}\left[B(g_1)B(g_2)\right] = \mathbf{E}\left[g_1(X_i)g_2(X_i)\right] - \mathbf{E}\left[g_1(X_i)\right]\mathbf{E}\left[g_2(X_i)\right].$$

*Then for any $\gamma \in (0,1)$, one can construct a random variable such that*

*(i)* $W_0 =_d \|B\|_{\mathcal{G}}$.

*(ii)*

$$\Pr\left(|W_n - W_0| > \frac{bK_n}{\gamma^{1/2}n^{1/2}} + \frac{\sigma^{1/2}K_n^{3/4}}{\gamma^{1/2}n^{1/2}} + \frac{b^{1/3}\sigma^{2/3}K_n^{2/3}}{\gamma^{1/2}n^{1/2}}\right) \leq A\left(\gamma + \frac{\log n}{n}\right),$$

*for some absolute constant $A$.*

Apply the above theorem with $\mathcal{G} = \Delta\mathcal{K}_n$, $\mathbb{G}_n(g) = Z(g)$ and $B(g) = G(g)$. We can check

$$\begin{aligned}
\mathbf{E}\left[G^2(g)\right] &= \mathbf{E}\left[Z^2(g)\right] \\
&= \mathbf{E}\left[np_n h_n^{J-1}\left(\Delta_n^{(j)}(y) - \mathbf{E}\left[\Delta_n^{(j)}(y)\right]\right)^2 / \sigma_{j,n}^2(y)\right] \\
&= 1.
\end{aligned}$$

Also, we obtain

$$\Pr\left(\left|\sup_{g\in\Delta\mathcal{K}_n} |Z(g)| - W_n^0\right| > \epsilon_{1n}\right) \leq \delta_{1n},$$

where $\epsilon_{1n} = \frac{b_nK_n}{\gamma^{1/2}n^{1/2}} + \frac{\sigma_n^{1/2}K_n^{3/4}}{\gamma^{1/2}n^{1/2}} + \frac{b_n^{1/3}\sigma_n^{2/3}K_n^{2/3}}{\gamma^{1/2}n^{1/2}}$ and $\delta_{1n} = A\left(\gamma + \frac{\log n}{n}\right)$. By letting $\gamma = Cn^{-c}$ for some $C$ and $0 < c < 1/2$, we have $\epsilon_{1n}$ and $\delta_{1n}$ bounded above by $C_1 n^{-c_1}$ for some $C_1$ and $c_1$.

To show $\mathbf{E}\left[\sup_{g\in\Delta\mathcal{K}_n} |G(g)|\right] \leq C_1\sqrt{\log n}$, we need Corollary 2.2.8 of Van der Vart and Wellner (2000).

*[Restatement of Corollary 2.2.8 of Van der Vart and Wellner (2000)]*

*Let $\{X_t : t \in T\}$ be a separable sub-Gaussian process with respect to the semimetric $d$. Then*

*(i) for every $\delta > 0$*

$$\mathbf{E}\left[\sup_{d(s,t)\leq\delta} |X_s - X_t|\right] \leq K\int_0^\delta \sqrt{\log D(\varepsilon, d)}d\varepsilon,$$

*for a university constant $K$, where $D(\varepsilon, d)$is the packing number for $T$. (The packing number is the maximum number of $\epsilon$-separated elements in $T$.)*

*(ii) In particular, for any $t_0$,*

$$\mathbf{E}\sup_t |X_t| \leq \mathbf{E}\,|X_{t_0}| + K\int_0^\infty \sqrt{\log D(\varepsilon, d)}d\varepsilon.$$

Applying (ii) of Corollary 2.2.8 to the Gaussian process $\{G(g) : g \in \Delta\mathcal{K}_n\}$ with respect to standard deviation semimetric $d$, we obtain

$$\mathbf{E}\left[\sup_{g\in\Delta\mathcal{K}_n} G(g)\right] \leq \mathbf{E}\,|G(g_0)| + K\int_0^\infty \sqrt{\log D(\varepsilon, d)}d\varepsilon.$$

By Condition SL1, $\int_0^\infty \sqrt{\log D(\varepsilon, d)}$ is finite and thus bounded by $C\log n$ for sufficiently large $n$.

**(ii) Verify SH2:** This follows Chernozhukov (2013) et al.
Define

$$\tilde{\mathbb{G}}_{n,j}(y) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i \frac{\sqrt{p_n h_n^{J-1}}\left[\varphi_n^{(j)}(Y_i, y) - \Delta_n^{(j)}(y)\right]}{\sigma_{j,n}(y)},$$

and

$$\Delta\mathbb{G}_{n,j}(y) = \hat{\mathbb{G}}_{n,j}(y) - \tilde{\mathbb{G}}_{n,j}(y).$$

Given any $y_1^n \in \mathbf{R}^{nJ}$, define

$$\begin{aligned}\hat{W}(y_1^n) &= \sup_{j\in\mathcal{J}, y\in\mathcal{Y}_{\kappa_2}} \hat{\mathbb{G}}_{n,j}(y),\\ \tilde{W}(y_1^n) &= \sup_{j\in\mathcal{J}, y\in\mathcal{Y}_{\kappa_2}} \tilde{\mathbb{G}}_{n,j}(y).\end{aligned}$$

Let $S_{n,1} \subset \mathbf{R}^{nJ}$ such that $\left|\frac{\hat{\sigma}_{j,n}(y)}{\sigma_{j,n}(y)} - 1\right| < \epsilon_{3n}$ for all $j$ and $y \in \mathcal{Y}_{\kappa_2}$ and for all $y_1^n \in S_{n,1}$, the "verify SH3" part below will show that $\Pr(S_{n,1}) > 1 - \delta_{3n} = 1 - 2/n$.

Fix any $y_1^n \in S_{n,1}$. Note that

$$\Delta\mathbb{G}_{n,j}(y) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i \frac{\sqrt{p_n h_n^{J-1}}\left[\varphi_n^{(j)}(y_1^n, y) - \Delta_n^{(j)}(y)\right]}{\sigma_{j,n}(y)}\left(\frac{\sigma_{j,n}(y)}{\hat{\sigma}_{j,n}(y)} - 1\right)$$

is a zero mean Gaussian process with variance

$$\begin{aligned}\mathbf{Var}\left(\Delta\mathbb{G}_{n,j}(y)\right) &= \frac{1}{n}\sum_{i=1}^n \frac{p_n h_n^{J-1}\left[\varphi_n^{(j)}(y_1^n, y) - \Delta_n^{(j)}(y)\right]^2}{\sigma_{j,n}^2(y)}\left(\frac{\sigma_{j,n}(y)}{\hat{\sigma}_{j,n}(y)} - 1\right)^2\\ &= \frac{\hat{\sigma}_{j,n}^2(y)}{\sigma_{j,n}^2(y)}\left(\frac{\sigma_{j,n}(y)}{\hat{\sigma}_{j,n}(y)} - 1\right)^2 \leq 4\epsilon_{3n}^2,\end{aligned}$$

where the last inequality comes from Condition SH3 and by using $\epsilon_{3n} \leq 1/2$, which holds for sufficiently large $n$.

In the following, we will need Proposition A.2.7 of Van der Vart and Wellner (2000).

*[Restatement of Proposition A.2.7 of Van der Vart and Wellner (2000)]*

*Let $X_t$, $t \in T$ be a separable zero-mean Gaussian process such that for some $K > \sigma(X)$, some $v > 0$ and some $0 < \varepsilon_0 \leq \sigma(X)$,*

$$N(\varepsilon, T, \rho) \leq (K/\varepsilon)^v \text{ for } 0 < \varepsilon < \varepsilon_0.$$

*Then there exists a universal constant $D$ such that for all $\lambda \geq \sigma^2(X)(1 + \sqrt{V})/\varepsilon_0$,*

$$\Pr\left(\sup_{t \in T} X_t \geq \lambda\right) \leq \left(\frac{DK\lambda}{\sqrt{V}\sigma^2(X)}\right)^v \frac{\sigma(X)}{\lambda} \exp(-\lambda^2/2\sigma^2(X)).$$

Now let $\epsilon_{3n} < 1/2$ and define

$$\Delta\tilde{\mathcal{K}}_n = \cup_{j=1}^J \Delta\tilde{\mathcal{K}}_{n,j} \text{ and } \Delta\tilde{\mathcal{K}}_{n,j} = \{ag : a \in (0,1], g \in \Delta\mathcal{K}_{n,j}\}.$$

By Condition SL1, the covering number of class $\Delta\tilde{\mathcal{K}}_n$ satisfies the polynomial bound condition of Proposition A.2.7 with $v > 1$, $\rho = L_2(Q)$, $T = \Delta\tilde{\mathcal{K}}_n$, and $\varepsilon_0 = 1$. Also, the uniform covering number of the Gaussian process $\Delta\mathbb{G}_{n,j}(y)$ with respect to the standard deviation semimetric is bounded by the uniform covering number of function $\Delta\tilde{\mathcal{K}}_n$. So the Gaussian process $\Delta\mathbb{G}_{n,j}(y)$ meets the conditions of Proposition A.2.7. Let $\lambda = K_n^{1/2}\epsilon_{3n}$ (where $K_n$ in specified in Condition SL4). Since $\sigma^2(X) = \mathbf{Var}(\Delta\mathbb{G}_{n,j}(y)) \leq 4\epsilon_{3n}^2$, we have $\lambda \geq \sigma^2(X)(1 + \sqrt{V})$. Applying Proposition A.2.7 to the process $\Delta\mathbb{G}_{n,j}(y)$, we obtain

$$\begin{aligned}
& \Pr\left(\left|\hat{W}(y_1^n) - \tilde{W}(y_1^n)\right| \geq \lambda\right) \\
\leq\; & \Pr\left(\sup_{j \in \mathcal{J}, y \in \mathcal{Y}_{n,\kappa_2}} \Delta\mathbb{G}_{n,j}(y) \geq \lambda\right) \\
\leq\; & \exp\left(\begin{array}{c} v\log(K_n^{3/2}/\mathbf{Var}(\Delta\mathbb{G}_{n,j}(y))) - 1/2\log K_n \\ -K_n\epsilon_{3n}^2/2\mathbf{Var}(\Delta\mathbb{G}_{n,j}(y)) \end{array}\right).
\end{aligned}$$

Since $\mathbf{Var}(\Delta\mathbb{G}_{n,j}(y)) \leq 4\epsilon_{3n}^2$, the leading term in the bracket is

$$-K_n\epsilon_{3n}^2/2\mathbf{Var}(\Delta\mathbb{G}_{n,j}(y)) \leq -K_n/8.$$

Recall that $K_n \geq \log n$ and $\lambda = K_n^{1/2}\epsilon_{3n}$ (can be bounded above by $Cn^{-c}$, for sufficiently large $n$, because $\epsilon_{3n} = 2\sqrt{b_n^2\sigma_n^2 K_n/n} \vee \sqrt{\sigma_n^2 K_n/n}$ and Condition SL4). Therefore, there exist constants $C_1$ and $c_1$ such that $\lambda \leq C_1 n^{-c_1}$ and

$$\Pr\left(\left|\hat{W}(y_1^n) - \tilde{W}(y_1^n)\right| \geq \lambda\right) < C_1 n^{-c_1}, \tag{A.11}$$

uniformly over $y_1^n \in S_{n,1}$.

We need Theorem 3.2 of Chernozhukov et al. (2013).

*[Restatement of Theorem 3.2 in Chernozhukov et al. (2013).*

*Suppose $\mathcal{G}$ is a class of measurable functions uniformly bounded by a constant $b$ such that there exist constants $a \geq e$ and $v > 1$ with $\sup_Q N(\mathcal{G}, L_2(Q), b\epsilon) \leq (a/\epsilon)^v$ for all $0 < \epsilon \leq 1$. Let $\sigma^2$ be a constant such that $\sup_{g\in\mathcal{G}} \mathbf{Var}(g) \leq \sigma^2 \leq b^2$. Assume that $b^2 K_n \leq n\sigma^2$, where $K_n$ specified in Theorem 3.1 cited before. Then for any $\delta > 0$, there exists a set $S_{n,0}$ such that $\Pr(S_{n,0}) \geq 1 - 3/n$ and for any $x_1^n \in S_{n,0}$ one can construct a random variable $W_0$ satisfying the following conditions.*

*(i) $W_0 =_d \|B\|_{\mathcal{G}}$ , where $B$ has the same property with that in Theorem 3.1 restated before.*

*(ii)*

$$\Pr\left(\left|\tilde{W}(x_1^n) - W_0\right| > \psi_n + \delta\right) \leq A\gamma_n(\delta),$$

*where $A$ is an absolute constant and*

$$\begin{aligned}\psi_n &= \left(\sigma^2 K_n/n\right)^{1/2} + \left(b^2\sigma^2 K_n^3/n\right)^{1/4},\\ \gamma_n(\delta) &= \left(b^2\sigma^2 K_n^3/n\right)^{1/4}/\delta + 1/n.\end{aligned}$$

Applying Theorem 3.2 to $\tilde{W}(y_1^n)$ and $B = G(g)$, we can construct a random variable $W^0 = \sup_{g\in\Delta\mathcal{K}_n} |G(g)|$ which satisfies the inequality in Theorem 3.2 (ii). Since $b_n^2 K_n^4/n \leq C_1 n^{-c_1}$ in Condition SL4, there exists a $\lambda$ such that $\lambda \leq C_1 n^{-c_1}$ and

$$\Pr\left(\left|\tilde{W}(y_1^n) - W_0\right| \geq \lambda\right) < C_1 n^{-c}, \tag{A.12}$$

uniformly over $y_1^n \in S_{n,2}$ and $\Pr(S_{n,2}) \geq 1 - 3/n$. Combining (A.11) and (A.12) leads to

$$\Pr\left(\left|\hat{W}(y_1^n) - W_0\right| \geq \lambda\right) < C_1 n^{-c}, \tag{A.13}$$

uniformly over $y_1^n \in S_n = S_{n,1} \cap S_{n,2}$, $\Pr(S_n) \geq 1 - 5/n$ and $\lambda \leq C_1 n^{-c_1}$.

Let $c^S_{1,n}(\alpha)$ be the conditional $(1-\alpha)$th quantile of the distribution of $\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{\kappa_2}}\left|\hat{\mathbb{G}}_{n,j}(y)\right|$ given $\{Y_i\}_{i=1}^n = y_1^n$. We obtain

$$\begin{aligned}
& \Pr\left(\sup_{g\in\Delta\mathcal{K}_n}|G(g)| \le c^S_{1,n}(\alpha)+\lambda\right) \\
= \;& \Pr\left(W^0 \le c^S_{1,n}(\alpha)+\lambda\right) \\
= \;& \Pr\left(\hat{W}(y_1^n) + \left(W^0 - \hat{W}(y_1^n)\right) \le c^S_{1,n}(\alpha)+\lambda\right) \\
\ge \;& \Pr\left(\hat{W}(y_1^n) \le c^S_{1,n}(\alpha)\right) - \Pr\left(\left|W^0 - \hat{W}(y_1^n)\right| > \lambda\right) \\
\ge \;& 1-\alpha - C_1 n^{-c_1},
\end{aligned}$$

where the first equality comes from $W^0 =_d \sup_{g\in\Delta\mathcal{K}_n}|G(g)|$, the third inequality comes from the fact that

$$\Pr(X+Y \le c+c') \ge \Pr(X\le c) - \Pr(Y>c'),$$

and the last inequality comes from the definition of $c^S_{1,n}(\alpha)$ and (A.13). Recall that $c_n$ be the $(1-\alpha)$th quantile of the distribution of $\sup_{g\in\Delta\mathcal{K}_n}|G(g)|$. We have

$$\Pr\left(\sup_{g\in\Delta\mathcal{K}_n}|G(g)| < c_n(\alpha + C_1 n^{-c_1})\right) = 1-\alpha - C_1 n^{-c_1}.$$

Therefore,

$$c^S_{1n}(\alpha) + \lambda \ge c_n(\alpha + C_1 n^{-c_1}),$$

uniformly over $y_1^n \in S_n$ such that $\Pr(S_n) \ge 1-5/n$ and $\lambda \le C_1 n^{-c_1}$. Thus Condition SH2 (i) holds with $\tau_n = C_1 n^{-c_1}$, $\epsilon_{2n} = \lambda \le C_1 n^{-c_1}$ and $\delta_{2n} = 5/n$. Condition SH2 (ii) can be proven analogously.

**(iii) Verify SH3**: This follows Chernozhukov et al. (2013)
Observe that

$$\left|\frac{\hat{\sigma}_{j,n}(y)}{\sigma_{j,n}(y)} - 1\right| \le \left|\frac{\hat{\sigma}^2_{j,n}(y)}{\sigma^2_{j,n}(y)} - 1\right|. \tag{A.14}$$

By the expressions of $\hat{\sigma}^2_{j,n}(y)$ and $\sigma^2_{j,n}(y)$ in Chapter 2,

$$\frac{\hat{\sigma}^2_{j,n}(y) - \sigma^2_{j,n}(y)}{\sigma^2_{j,n}(y)} = \mathbf{E}_n\left[\left(\frac{\varphi_n^{(j)}(Y_i,y)}{\sigma_{j,n}(y)}\right)^2\right] - \mathbf{E}\left[\left(\frac{\varphi_n^{(j)}(Y_i,y)}{\sigma_{j,n}(y)}\right)^2\right]$$

$$+\left(\mathbf{E}_n\left[\frac{\varphi_n^{(j)}(Y_i,y)}{\sigma_{j,n}(y)}\right]\right)^2-\left(\mathbf{E}\left[\frac{\varphi_n^{(j)}(Y_i,y)}{\sigma_{j,n}(y)}\right]\right)^2,$$

where $\mathbf{E}_n$ denotes the sample mean. Therefore,

$$\begin{aligned}\left|\frac{\hat{\sigma}^2_{j,n}(y)}{\sigma^2_{j,n}(y)}-1\right| &\leq \sup_{g\in\Delta\mathcal{K}_n^2}\left|\mathbf{E}_n\left[g(Y_i)^2\right]-\mathbf{E}\left[g(Y_1)^2\right]\right| \\ &\quad+\sup_{g\in\Delta\mathcal{K}_n}\left|\left(\mathbf{E}_n\left[g(Y_1)\right]\right)^2-\left(\mathbf{E}\left[g(Y_1)\right]\right)^2\right|, \qquad \text{(A.15)}\end{aligned}$$

where $\Delta\mathcal{K}_n^2=\cup_{j=1}^J\Delta\mathcal{K}_{n,j}^2$ and $\Delta\mathcal{K}_n^2=\left\{g^2:g\in\Delta\mathcal{K}_{n,j}\right\}$. By the definition of $\sigma_n^2$ in Condition SL4, $g\leq b_n^2$. Thus for all $g\in\Delta\mathcal{K}_n^2$,

$$\mathbf{E}\left[g(Y_1)^2\right]\leq b_n^2\mathbf{E}\left[g(Y_1)\right]\leq b_n^2\sigma_n^2,$$

where the last inequality comes from that $\mathbf{E}\left[g(Y_i)\right]\leq\sigma_n^2$ for $g\in\Delta\mathcal{K}_n^2$, by Condition SL1(ii) and the definition of $\sigma_n^2$ in Condition SL4. In addition, the covering number of $\mathcal{K}_n^2$ also satisfies the polynomial bound (A.10). In the following, we will use a form of Talagrand's Inequality stated and proven by Chernozhukov et al. (2013).

*[Restatement of Theorem A.4 in Chernozhukov et al. (2013)]*

*Let $\xi_1$,...,$\xi_n$ be i.i.d. random variables taking values in a measurable space $(S,\mathcal{S})$. Suppose that $\mathcal{G}$ is a non-empty, pointwise measurable class of functions on $S$ uniformly bounded by a constant $b$ such that there exist constants $a\geq e$ and $v>1$ with $\sup_Q N(\mathcal{G},L_2(Q),b\epsilon)\leq(a/\epsilon)^v$ for all $0<\epsilon\leq1$. Let $\sigma^2$ be a constant such that $\sup_{g\in\mathcal{G}}\mathbf{var}(g)\leq\sigma^2\leq b^2$.If $b^2v\log(ab/\sigma)\leq n\sigma^2$, then for all $t\leq n\sigma^2/b^2$,*

$$\Pr\left[\sup_{g\in\mathcal{G}}\left|\sum_{i=1}^n\left(g(\xi_i)-\mathbf{E}\left[g(\xi_1)\right]\right)\right|>A\sqrt{n\sigma^2\left\{t\vee v\log(ab/\sigma)\right\}}\right]\leq e^{-t},$$

*where $A>0$ is an absolute constant.*

Let $t=\log n$, $b=b_n$ and $\sigma=\sigma_n$. (Note that $t\leq n\sigma_n^2/b_n^2$ for large $n$.) Recall that $K_n=Av(\log n\vee\log(ab_n/\sigma_n))$. We have

$$b_n^2K_n/n\leq b_n^2K_n^4/n\leq C_5n^{-c_5}\leq c_2\leq\sigma_n^2,$$

where the first and third inequality hold for sufficiently large $n$, the second inequality follows from Condition SL4 and the last inequality follows from Condition SL2. Since $K_n\geq Av\log(ab_n/\sigma_n)$, $b_n^2v\log(ab_n/\sigma_n)\leq\sigma_n^2$ and

Condition SL1 holds, conditions of Theorem A.4 are satisfied. Applying Theorem A.4 to $\sup_{g\in\Delta\mathcal{K}_n^2}\left|\mathbf{E}_n\left[g(Y_i)^2\right]-\mathbf{E}\left[g(Y_1)^2\right]\right|$, we have

$$\Pr\left[\begin{array}{l}\sup_{g\in\Delta\mathcal{K}_n^2}\left|\mathbf{E}_n\left[g(Y_i)^2\right]-\mathbf{E}\left[g(Y_1)^2\right]\right| \\ > A\sqrt{n\sigma_n^2\left\{\log n\vee v\log(ab_n/\sigma_n)\right\}/n}\end{array}\right]\leq 1/n.$$

Notice that $\log n\vee v\log(ab_n/\sigma_n)\leq v\left\{\log n\vee\log(ab_n/\sigma_n)\right\}\leq K_n$, where the first inequality comes from $v>1$ and the second from any $A\geq 1$. We obtain

$$\Pr\left[\sup_{g\in\Delta\mathcal{K}_n^2}\left|\mathbf{E}_n\left[g(Y_i)^2\right]-\mathbf{E}\left[g(Y_1)^2\right]\right|>\sqrt{\sigma_n^2K_n/n}\right]\leq 1/n. \tag{A.16}$$

The second term on the right hand side of (A.15) is bounded above by

$$\begin{aligned}&\sup_{g\in\Delta\mathcal{K}_n}\left|\left(\mathbf{E}_n\left[g(Y_i)\right]\right)^2-\left(\mathbf{E}\left[g(Y_1)\right]\right)^2\right| \\ \leq\ & 2b_n\sup_{g\in\Delta\mathcal{K}_n}\left|\mathbf{E}_n\left[g(Y_i)^2\right]-\mathbf{E}\left[g(Y_1)^2\right]\right|,\end{aligned}$$

where the inequality comes from Condition SL1(ii).

Again applying Theorem A.4 to $\sup_{g\in\Delta\mathcal{K}_n}\left|\mathbf{E}_n\left[g(Y_i)\right]-\mathbf{E}\left[g(Y_1)\right]\right|$, we have

$$\Pr\left[\sup_{g\in\Delta\mathcal{K}_n}\left|\mathbf{E}_n\left[g(Y_i)\right]-\mathbf{E}\left[g(Y_1)\right]\right|>\sqrt{\sigma_n^2K_n/n}\right]\leq 1/n.$$

Thus

$$\Pr\left[\sup_{g\in\Delta\mathcal{K}_n}\left|\left(\mathbf{E}_n\left[g(Y_1)\right]\right)^2-\left(\mathbf{E}\left[g(Y_1)\right]\right)^2\right|>2\sqrt{b_n^2\sigma_n^2K_n/n}\right]\leq 1/n. \tag{A.17}$$

Let $\epsilon_{3n}=2\sqrt{b_n^2\sigma_n^2K_n/n}\vee\sqrt{\sigma_n^2K_n/n}$. Then there exist constants $C_1$ and $c_1$ such that $\epsilon_{3n}\leq C_1n^{-c_1}$. By letting $\delta_{3n}=2/n$, Condition SH3 follows.

**(iv) Verify SH4**: By Condition SH3 and SL2, we know that with probability $1-\delta_{3n}$. Thus we have

$$\begin{aligned}c_{1,n}^S(\gamma_n)\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{\kappa_2}}\hat{\sigma}_{j,n}(y) &\leq c_{1,n}^S(\gamma_n)(1+\epsilon_{3n})\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{\kappa_2}}\hat{\sigma}_{j,n,s}(y) \\ &\leq Cc_{1,n}^S(\gamma_n),\end{aligned}$$

where $C\geq(1+\epsilon_{3n})C_2$. Combining this with Condition SL4, we have

$$\sqrt{np_n h_n^{J-1}} p_n^t \leq \bar{M} C c_{1,n}^S(\gamma_n). \tag{A.18}$$

Observe that

$$\begin{aligned}
B_n &\leq \frac{\sqrt{np_n h_n^{J-1}} C_4 p_n^t}{\inf_{j \in \mathcal{J}, y \in \mathcal{Y}_{\kappa_2}} \hat{\sigma}_{j,n}(y)} \\
&\leq \frac{\sqrt{np_n h_n^{J-1}} C_4 p_n^t}{(1-\epsilon_{3n}) \inf_{j \in \mathcal{J}, y \in \mathcal{Y}_{\kappa_2}} \sigma_{j,n}(y)} \\
&\leq \frac{\sqrt{np_n h_n^{J-1}} C_4 p_n^t}{(1-\epsilon_{3n}) c_2} \leq C_6 \sqrt{np_n h_n^{J-1}} p_n^t \\
&\leq C_6 \bar{M} C c_{1,n}^S(\gamma_n),
\end{aligned}$$

where the first inequality comes from Condition SL3, the second inequality comes from Condition SH3, the third follows from Condition SL2, the fourth holds for any constant $C_6$ satisfying $C_6 \geq C_4/((1-\epsilon_{3n})c_2)$, the last comes from (A.18). **Q.E.D.**

### Lemmas

The following three lemmas will be used in the proof of Proposition 5 and 6. Lemma A1(i) and Lemma A3 will be used in the proof of Proposition 5. Lemma A1(ii) will be used in the proof of Proposition 6. Lemma 2 will be used in the proof of Lemma 3.

**Lemma A1**. *(i) Suppose that $H_0$ is true, or Condition A1 holds. If Condition SL5 holds,*

$$\Pr\left(\hat{\mathcal{Y}}_{n,\kappa_1} \subseteq \mathcal{Y}_{\kappa_2}\right) \geq 1 - O(n^{-c_0}),$$

*for some $c_0 > 0$.*

*(ii) Suppose $H_0$ is true, or Condition A1 holds. If Condition SL5 holds and*

$$0 < d_n < \min\left(\gamma_h, \gamma_p, (1-\gamma_p-(J-1)\gamma_h)/2\right),$$

*then*

$$\Pr\left(\mathcal{Y}_{n,L,c} \subseteq \hat{\mathcal{Y}}_{n,\kappa_1}\right) \geq 1 - O(n^{-c_0'}),$$

*for some* $c_0' > 0$.

**Proof of Lemma A1.**
(i) Define

$$\begin{aligned}
\tilde{P}_j^-(y) &\equiv \frac{1}{n}\sum_{i=1}^{n} 1\{y_j - w_j \le Y_{i,j} < y_j, |Y_{i,k} - y_k| \le h_n/2, \text{all } k \ne j\}, \\
\tilde{P}_j^+(y) &\equiv \frac{1}{n}\sum_{i=1}^{n} 1\{y_j \le Y_{i,j} \le y_j + w_j, |Y_{i,k} - y_k| \le h_n/2, \text{all } k \ne j\}, \\
\tilde{P}_j(y) &= \tilde{P}_j^-(y) + \tilde{P}_j^+(y), \;\; w_j = (\sup_y Y_j - \inf_y Y_j)/H,
\end{aligned}$$

and

$$\tilde{\mathcal{Y}}_{\kappa_1} = \{y \in \mathcal{Y} : \tilde{P}_j^-(y)/\tilde{P}_j(y) > \kappa_1, \tilde{P}_j^+(y)/\tilde{P}_j(y) > \kappa_1, \text{ all } j\}.$$

We know that $\hat{\mathcal{Y}}_{n,\kappa_1} \subseteq \tilde{\mathcal{Y}}_{\kappa_1}$ for all $n$ since $w_j \ge \hat{w}_j$. It suffices to show that

$$\Pr\left(\tilde{\mathcal{Y}}_{\kappa_1} \subseteq \mathcal{Y}_{\kappa_2}\right) \ge 1 - O(n^{-c_0}).$$

By definitions of $\tilde{\mathcal{Y}}_{\kappa_1}$, $\mathcal{Y}_{\kappa_2}$ and Markov's inequality,

$$\begin{aligned}
\Pr\left(\tilde{\mathcal{Y}}_{\kappa_1} \subseteq \mathcal{Y}_{\kappa_2}\right) &\ge 1 - \sum_{j=1}^{J} \Pr\left(\left|\tilde{P}_j^-(y)/\tilde{P}_j(y) - P_j^-(y)/P_j(y)\right| > (\kappa_1 - \kappa_2)\right) \\
&\quad - \sum_{j=1}^{J} \Pr\left(\left|\tilde{P}_j^+(y)/\tilde{P}_j(y) - P_j^+(y)/P_j(y)\right| > (\kappa_1 - \kappa_2)\right) \\
&\ge 1 - \sum_{j=1}^{J} \mathbf{E}\left[\left(\tilde{P}_j^-(y)/\tilde{P}_j(y) - P_j^-(y)/P_j(y)\right)^2\right] / (\kappa_1 - \kappa_2)^2 \\
&\quad - \sum_{j=1}^{J} \mathbf{E}\left[\left(\tilde{P}_j^+(y)/\tilde{P}_j(y) - P_j^+(y)/P_j(y)\right)^2\right] / (\kappa_1 - \kappa_2)^2 \\
&= \begin{cases} 1 - O\left(p_n^4 + h_n^4 + \frac{1}{np_n h_n^{J-1}}\right) & \text{under } H_0, \\ 1 - O\left(p_n^2 + h_n^2 + \frac{1}{np_n h_n^{J-1}}\right) & \text{under Condition A1,} \end{cases} \\
&= 1 - O(n^{-c_0}).
\end{aligned}$$

The last equality is from Condition SL5, for a properly chosen $c_0$ depending on the rates $\gamma_h$ and $\gamma_p$.

Part (ii) can be proven in the same way. We have

$$\Pr\left(\mathcal{Y}_{n,L,c} \subseteq \hat{\mathcal{Y}}_{n,\kappa_1}\right) \quad \geq \quad \begin{cases} 1 - O\left((p_n^4 + h_n^4 + \frac{1}{np_n h_n^{J-1}})n^{2d}\right) & \text{under } H_0, \\ 1 - O\left((p_n^2 + h_n^2 + \frac{1}{np_n h_n^{J-1}})n^{2d}\right) & \text{under Condition A1,} \end{cases}$$

The desired result follows from the restriction imposed on $d$ by Condition (ii) of Lemma A1. **Q.E.D.**

The following Condition A2 will be used in Lemma A2.

**Condition A2**. The alternative hypothesis $H_1$ is true, and the density $f_Y(y) > 0$ for all $y \in \mathcal{Y}_{\kappa_1}$, where

$$\mathcal{Y}_{\kappa_1} = \{y \in \mathcal{Y} : P_j^-(y)/P_j(y) > \kappa_1, P_j^+(y)/P_j(y) > \kappa_2, \text{ all } j \in \mathcal{J}\}.$$

**Lemma A2**. *Suppose $H_0$ is true or Conditions A1 and A2 hold. If Conditions SL2-SL5 hold, given any $y_1^n \in S_n$, where $S_n$ is defined when verifying Condition SH2 and $\Pr(S_n) > 1 - 5/n$, we have*

$$\Pr\left(\left|\sup_{j\in\mathcal{J}, y\in\hat{\mathcal{Y}}_{n,\kappa_2}} \hat{\mathbb{G}}_{n,j}(y) - \sup_{j\in\mathcal{J}, y\in\mathcal{Y}_{\kappa_2}} \hat{\mathbb{G}}_{n,j}(y)\right| > r_n\right) \leq \nu_n,$$

*for sufficiently large $n$, where $r_n$ and $\nu_n$ are bounded above by some $Cn^{-c}$.*

**Proof of Lemma A2.**
Let

$$\begin{aligned} \mathcal{Y}_{L,n,\kappa_2} &= \{y \in \mathcal{Y} : \frac{P_j^-(y)}{P_j(y)} > \kappa_2 + n^{-d}, \frac{P_j^+(y)}{P_j(y)} > \kappa_2 + n^{-d}, \text{all } j\}, \\ \mathcal{Y}_{U,n,\kappa_2} &= \{y \in \mathcal{Y} : \frac{P_j^-(y)}{P_j(y)} > \kappa_2 - n^{-d}, \frac{P_j^+(y)}{P_j(y)} > \kappa_2 - n^{-d}, \text{all } j\}, \end{aligned}$$

where $d < \min(\gamma_p, \gamma_h, (1 - \gamma_p - (J-1)\gamma_h)/2)$. By similar arguments as in the proof of Lemma A1, we have

$$\Pr\left(\mathcal{Y}_{L,n,\kappa_2} \subset \hat{\mathcal{Y}}_{n,\kappa_2} \subset \mathcal{Y}_{U,n,\kappa_2}\right) \geq 1 - O(n^{-k}),$$

where $k = \min(\gamma_p, \gamma_h, (1 - \gamma_p - (J-1)\gamma_h)/2) - d$. By definition, $\mathcal{Y}_{L,n,\kappa_2} \subset \mathcal{Y}_{\kappa_2} \subset \mathcal{Y}_{U,n,\kappa_2}$. Therefore, for some $r_n \to 0$ ($r_n$ will be specified later),

$$\begin{aligned}
& \Pr\left(\left|\sup_{j\in\mathcal{J},y\in\hat{\mathcal{Y}}_{n,\kappa_2}}\hat{\mathbb{G}}_{n,j}(y)-\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{\kappa_2}}\hat{\mathbb{G}}_{n,j}(y)\right|>r_n\right)\\
\leq\ & \Pr\left(\left|\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{L,n,\kappa_2}}\hat{\mathbb{G}}_{n,j}(y)-\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{U,n,\kappa_2}}\hat{\mathbb{G}}_{n,j}(y)\right|>r_n\right)\\
& +O(n^{-k}).
\end{aligned} \tag{A.19}$$

By (A.11) (with a slight modification of the index set), we have

$$\begin{aligned}
\Pr\left(\left|\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{L,n,\kappa_2}}\hat{\mathbb{G}}_{n,j}(y)-\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{L,n,\kappa_2}}\tilde{\mathbb{G}}_{n,j}(y)\right|\geq r_n/3\right) &< C_1n^{-c_1},\text{ and}\\
\Pr\left(\left|\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{U,n,\kappa_2}}\hat{\mathbb{G}}_{n,j}(y)-\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{U,n,\kappa_2}}\tilde{\mathbb{G}}_{n,j}(y)\right|\geq r_n/3\right) &< C_1n^{-c_1},
\end{aligned}$$

uniformly over $y_1^n\in S_n$ and $r_n/3\leq C_1n^{-c_1}$. Given $y_1^n\in S_n$, we have

$$\begin{aligned}
& \Pr\left(\left|\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{L,n,\kappa_2}}\hat{\mathbb{G}}_{n,j}(y)-\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{U,n,\kappa_2}}\hat{\mathbb{G}}_{n,j}(y)\right|>r_n\right)\\
\leq\ & \Pr\left(\left|\max_{j,y\in y\in\mathcal{Y}_{L,n,\kappa_2}}\tilde{\mathbb{G}}_{n,j}(y)-\max_{j,y\in\mathcal{Y}_{U,n,\kappa_2}}\tilde{\mathbb{G}}_{n,j}(y)\right|\geq r_n/3\right)\\
& +\Pr\left(\left|\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{U,n,\kappa_2}}\hat{\mathbb{G}}_{n,j}(y)-\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{U,n,\kappa_2}}\tilde{\mathbb{G}}_{n,j}(y)\right|\geq r_n/3\right)\\
& +\Pr\left(\left|\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{L,n,\kappa_2}}\hat{\mathbb{G}}_{n,j}(y)-\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{L,n,\kappa_2}}\tilde{\mathbb{G}}_{n,j}(y)\right|\geq r_n/3\right)\\
\leq\ & \Pr\left(\left|\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{L,n,\kappa_2}}\tilde{\mathbb{G}}_{n,j}(y)-\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{U,n,\kappa_2}}\tilde{\mathbb{G}}_{n,j}(y)\right|\geq r_n/3\right)\\
& +C_1n^{-c_1}.
\end{aligned} \tag{A.20}$$

By Condition A2, for any $y'\in\mathcal{Y}_{U,n,\kappa_2}\backslash\mathcal{Y}_{L,n,\kappa_2}$, there exists some $y''\in\mathcal{Y}_{L,n,\kappa_2}$ such that $|y'-y''|<\epsilon$, for any $\epsilon$, when $n$ is sufficiently large. Hence

$$\begin{aligned}
& \Pr\left(\left|\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{U,n,\kappa_2}}\tilde{\mathbb{G}}_{n,j}(y)-\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{L,n,\kappa_2}}\tilde{\mathbb{G}}_{n,j}(y)\right|\geq r_n/3\right)\\
\leq\ & \Pr\left(\sup_{j\in\mathcal{J},|y_1-y_0|<\epsilon,y_1,y_0\in\mathcal{Y}_{U,n,\kappa_2}}\left|\tilde{\mathbb{G}}_{n,j}(y_0)-\tilde{\mathbb{G}}_{n,j}(y_1)\right|\geq r_n/3\right).
\end{aligned}$$

We can write the process $\{\tilde{\mathbb{G}}_{n,j}(y) : y \in \mathcal{Y}_{U,n,\kappa_2}, j \in \mathcal{J}\}$ as $\tilde{\mathbb{G}} = \{\tilde{\mathbb{G}}(g) : g \in \Delta\mathcal{K}_{U,n}\}$, where

$$\tilde{\mathbb{G}}(g) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\left[g(Y_i) - \frac{1}{n}\sum_{i=1}^{n}g(Y_i)\right], g \in \Delta\mathcal{K}_{U,n},$$

and $\mathcal{K}_{U,n} = \cup_{j=1}^{J}\mathcal{K}_{U,n,j}$, where

$$\Delta\mathcal{K}_{U,n,j} = \left\{\frac{\sqrt{p_n h_n^{J-1}}\varphi_n^{(j)}(\cdot, y)}{\sigma_{n,j}(y)} : y \in \mathcal{Y}_{U,n,\kappa_2}\right\}.$$

For fixed $y_1^n$, $\tilde{\mathbb{G}}$ is a centred Gaussian process and is sub-Gaussian with respect to standard deviation semimetric $\sigma$ such that $\sigma(g_1, g_0) = (\mathbf{E}[\tilde{\mathbb{G}}(g_1) - \tilde{\mathbb{G}}(g_2)]^2)^{1/2}$ . In addition, by the mean square continuous sample path of $\tilde{\mathbb{G}}$, it must have $\sigma(g_1, g_0) \to 0$ as $|y_1 - y_0| \to 0$, where $g_1, g_0 \in \Delta\mathcal{K}_{U,n,j}$, $y_1, y_0 \in \mathcal{Y}_{U,n,\kappa 2}$, and

$$g_\lambda : t \longmapsto \frac{\sqrt{p_n h_n^{J-1}}\varphi_n^{(j)}(t, y_\lambda)}{\sigma_{n,j}(y_\lambda)},\ \lambda = 0, 1.$$

Therefore, for every $\delta > 0$, we obtain

$$\begin{aligned}
&\Pr\left(\sup_{j\in\mathcal{J},|y_1-y_0|<\epsilon,y_1,y_0\in\mathcal{Y}_{U,n,\kappa_2}}\left|\tilde{\mathbb{G}}_{n,j}(y_0) - \tilde{\mathbb{G}}_{n,j}(y_1)\right| \geq r_n/3\right) \\
\leq\ &\Pr\left(\sup_{\sigma(g_1,g_0)<\delta,g_0,g_1\in\Delta\mathcal{K}_{U,n}}\left|\tilde{\mathbb{G}}(g_1) - \tilde{\mathbb{G}}(g_0)\right| \geq r_n/3\right) \\
\leq\ &3\mathbf{E}\left[\sup_{\sigma(g_1,g_0)<\delta,g_0,g_1\in\Delta\mathcal{K}_{U,n}}\left|\tilde{\mathbb{G}}(g_1) - \tilde{\mathbb{G}}(g_0)\right|\right]/r_n \\
\leq\ &\frac{3K}{r_n}\int_0^\delta\sqrt{\log D(\varepsilon, \sigma)}d\varepsilon. \qquad\qquad (A.21)
\end{aligned}$$

The second inequality comes from Markov's inequality and the last from Corollary 2.2.8 of Van der Vart and Wellner (2000). $D(\varepsilon, \sigma)$ is the packing number of the class $\Delta\mathcal{K}_n$ with radius $\varepsilon$ and standard deviation semimetric. By Condition SL1 and the fact that $D(\varepsilon, \sigma) < N(\varepsilon/2, \sigma)$, we have $\log D(\varepsilon, \sigma) < K_1 \log(1/\varepsilon)$ for some constant $K_1$. For sufficiently small $\delta$, we obtain

$$\int_0^\delta\sqrt{\log D(\varepsilon, \sigma)}d\varepsilon \leq K_2\int_0^\delta\log(1/\varepsilon)d\varepsilon \leq K_2(\delta(\log(1/\delta + 1))),$$

for some constant $K_2$. Now let $\delta = 1/n$ and $r_n = n^{-r}$. We have

$$\frac{n^{c_1}}{r_n}\int_0^{\delta}\sqrt{\log D(\varepsilon,\sigma)}d\varepsilon \le \frac{\log(n+1)}{n^{1-c_1-r_1}}.$$

The right hand side of the above inequality goes to zero when $n \to \infty$, as long as $c_1 + r < 1$. As a result, $\frac{3K}{r_n}\int_0^{\delta}\sqrt{\log D(\varepsilon,\sigma)}d\varepsilon < C_1 n^{-c_1}$ for sufficiently large $n$.
Lemma A2 follows by combining (A.19), (A.20) and (A.21) and by letting $\nu_n$ equal some some constant times $n^{-(c_1 \wedge k)}$. **Q.E.D.**

**Lemma A3**. *Suppose Conditions SL1-SL5 hold, there exist numbers $c$ and $C$ satisfying*

$$\Pr\left(\hat{c}^S_{1,n}(\alpha) \le c^S_{1,n}(\alpha+\nu_n) - r_n\right) \le \delta_{5n},$$

*for some positive sequence $r_n$, $v_n$ and $\delta_{5n}$ bounded above by $Cn^{-c}$.*

**Proof of Lemma A3.**
For any $y_1^n \in S_n$ (recall $\Pr(S_n) > 1 - 5/n$),

$$\begin{aligned}
&\Pr\left(\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{\kappa_2}}\left|\hat{\mathbb{G}}_{n,j}(y)\right| \le \hat{c}^S_{1,n}(\alpha)+r_n\right)\\
\ge\ &\Pr\left(\begin{array}{c}\left|\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{\kappa_2}}\left|\hat{\mathbb{G}}_{n,j}(y)\right| - \sup_{j\in\mathcal{J},y\in\hat{\mathcal{Y}}_{n,\kappa_2}}\left|\hat{\mathbb{G}}_{n,j}(y)\right|\right|\\ +\sup_{j\in\mathcal{J},y\in\hat{\mathcal{Y}}_{n,\kappa_2}}\left|\hat{\mathbb{G}}_{n,j}(y)\right| \le \hat{c}^S_{1,n}(\alpha)+r_n\end{array}\right)\\
\ge\ &\Pr\left(\sup_{j\in\mathcal{J},y\in\hat{\mathcal{Y}}_{n,\kappa_2}}\left|\hat{\mathbb{G}}_{n,j}(y)\right| \le \hat{c}^S_{1,n}(\alpha)\right)\\
&-\Pr\left(\left|\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{\kappa_2}}\left|\hat{\mathbb{G}}_{n,j}(y)\right| - \sup_{j\in\mathcal{J},y\in\hat{\mathcal{Y}}_{n,\kappa_2}}\left|\hat{\mathbb{G}}_{n,j}(y)\right|\right| > r_n\right)\\
\ge\ &1-\alpha-\nu_n,
\end{aligned}$$

where the last inequality comes from the definition of $\hat{c}^S_{1,n}(\alpha)$ and Lemma A2. Meanwhile, by definition of $c^S_{1,n}(\cdot)$,

$$\Pr\left(\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{\kappa_2}}\left|\hat{\mathbb{G}}_{n,j}(y)\right| \le c^S_{1,n}(\alpha+\nu_n)\right) = 1-\alpha-\nu_n.$$

Thus we obtain

$$\hat{c}^S_{1,n}(\alpha)+r_n \ge c^S_{1,n}(\alpha+\nu_n),$$

for any $y_1^n \in S_n$. Lemma A3 follows by letting $\delta_{5n} = 5/n$. **Q.E.D.**

**Proofs of Proposition 5 and 6**

We prove Proposition 5 and 6 using Conditions SL1-SL4 and Lemmas A1 and A3.

**Proof of Proposition 5.**

Observe that

$$
\begin{aligned}
& \Pr\left(\Delta_n < \tilde{c}_{1,n}^S\right) \\
\geq\ & _{(1)} \Pr\left(\sup_{j\in\mathcal{J}, y\in\mathcal{Y}_{\kappa_2}} \left|\frac{\sqrt{np_n h_n^{J-1}}\Delta_n^{(j)}(y)}{\hat{\sigma}_{j,n}(y)}\right| < \tilde{c}_{1,n}^S\right) - \mu_n \\
\geq\ & _{(2)} \Pr\left(\sup_{j\in\mathcal{J}, y\in\mathcal{Y}_{\kappa_2}} |Z_{j,n}(y)| \frac{\sigma_{j,n}(y)}{\hat{\sigma}_{j,n}(y)} + B_n < \tilde{c}_{1,n}^S\right) - \mu_n \\
\geq\ & _{(3)} \Pr\left(\sup_{j\in\mathcal{J}, y\in\mathcal{Y}_{\kappa_2}} |Z_{j,n}(y)| \frac{\sigma_{j,n}(y)}{\hat{\sigma}_{j,n}(y)} + B_n < c_{1,n}^S(\nu_n) - r_n\right) - \mu_n - \delta_{5n} \\
\geq\ & _{(4)} \Pr\left(\sup_{j\in\mathcal{J}, y\in\mathcal{Y}_{\kappa_2}} |Z_{j,n}(y)| \frac{\sigma_{j,n}(y)}{\hat{\sigma}_{j,n}(y)} < c_{1,n}^S(\alpha+\nu_n) - r_n\right) \\
& -\mu_n - \delta_{5n} - \delta_{4n} \\
\geq\ & _{(5)} \Pr\left(\begin{array}{c} \sup_{g\in\Delta\mathcal{K}_n} |Z_n(g)| < (1-\epsilon_{3n})c_n(\alpha+\tau_n+\nu_n) \\ -(1-\epsilon_{3n})r_n - (1-\epsilon_{3n})\epsilon_{2n} - (1-\epsilon_{3n})r_n \end{array}\right) \\
& -\mu_n - \delta_{5n} - \delta_{4n} - \delta_{2n} - \delta_{3n} \\
\geq\ & _{(6)} \Pr\left(\begin{array}{c} \sup_{g\in\Delta\mathcal{K}_n} |G(g)| < (1-\epsilon_{3n})c_n(\alpha+\tau_n+\nu_n) \\ -(1-\epsilon_{3n})r_n - \epsilon_{1n} - (1-\epsilon_{3n})\epsilon_{2n} \end{array}\right) \\
& -\mu_n - \delta_n \\
\geq\ & _{(7)} \Pr\left(\sup_{g\in\Delta\mathcal{K}_n} |G(g)| < c_n(\alpha+\tau_n+\nu_n)\right) \\
& - \sup_{w\in\mathbf{R}} \Pr\left(\left|\sup_{g\in\Delta\mathcal{K}_n} |G(g)| - w\right| < \epsilon_n\right) - \mu_n - \delta_n \\
\geq\ & _{(8)} 1 - \alpha - \tau_n - \nu_n - A\epsilon_n \mathbf{E}\left|\sup_{g\in\Delta\mathcal{K}_n} |G(g)|\right| - \mu_n - \delta_n \\
\geq\ & _{(9)} 1 - \alpha - Cn^{-c},
\end{aligned}
$$

where $\delta_n = \delta_{1n} + \delta_{2n} + \delta_{3n} + \delta_{4n} + \delta_{5n}$, and $\epsilon_n = c_n(\alpha+\tau_n+\nu_n)\epsilon_{3n} + \epsilon_{1n} +$

$(1-\epsilon_{3n})\epsilon_{2n}+(1-\epsilon_{3n})r_n$. Inequality (1) comes from Lemma A1. Inequality (2) comes from

$$\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{\kappa_2}}\left|\frac{\sqrt{np_nh_n^{J-1}}\Delta_n^{(j)}(y)}{\hat{\sigma}_{j,n}(y)}\right|\leq\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{\kappa_2}}|Z_{j,n}(y)|\frac{\sigma_{j,n}(y)}{\hat{\sigma}_{j,n}(y)}+B_n,$$

under $H_0$. (This is because $\Delta_n(y)=0$ for all $j$ under $H_0$.) Inequality (3) comes from Lemma A3. Inequality (4) follows $c_{1,n}\equiv c_{1,n}(\alpha)+c'_{1,n}$, $\Pr(X+Y\leq c+c')\geq\Pr(X\leq c)-\Pr(Y>c')$ and Condition SH4. Inequality (5) comes from Condition SH2, SH3 and the fact that $\Pr(X<Y)\geq\Pr(X<a)-\Pr(Y<a)$. Inequality (6) comes from Condition SH1(ii). Inequality (7) uses the fact that

$\Pr(X<a)-\Pr(X<a-b)\leq\Pr(|X-x|<b)$ for $a-b<x<a, a>0, b>0$.

Inequality (8) follows by applying Corollary 2.1 of Chernozhukov et al.(2013) to the second term on the right hand side of inequality (7). Inequality (9) uses Condition SH1(i) together with Markov's inequality to bound $c_{1,n}(\alpha)$ and $\sup_{g\in\Delta\mathcal{K}_n}|G(g)|$. The argument from inequality (4) to the end are adapted from the proof of Proposition 4.1 in Chernozhukov et al.(2013). **Q.E.D.**

**Proof of Proposition 6.** Recall the (2.19) and let

$$B_{j,n}(y)=\frac{\sqrt{np_nh_n^{J-1}}\left(\mathbf{E}\left[\Delta_n^{(j)}(y)\right]-\Delta^{(j)}(y)\right)}{\hat{\sigma}_{j,n}(y)}.$$

Notice that

$$\begin{aligned}&\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{\kappa_2}}\left|\frac{\sqrt{np_nh_n^{J-1}}\Delta_n^{(j)}(y)}{\hat{\sigma}_{j,n}(y)}\right|\\ =&\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{\kappa_2}}\left||Z_{j,n}(y)|\frac{\sigma_{j,n}(y)}{\hat{\sigma}_{j,n}(y)}+B_{j,n}(y)+\frac{\sqrt{np_nh_n^{J-1}}\Delta^{(j)}(y)}{\hat{\sigma}_{j,n}(y)}\right|.\end{aligned}$$

Therefore,

$$\Pr\left(\Delta_n>\tilde{c}_{1,n}^S\right)$$

$$
\begin{aligned}
&\geq \Pr\left(\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{n,L,c}}\left|\frac{\sqrt{np_nh_n^{J-1}}\Delta_n^{(j)}(y)}{\hat{\sigma}_{j,n}(y)}\right|>\tilde{c}_{1,n}^S\right)-o(1)\\
&= \Pr\left(\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{n,L,c}}\left||Z_{j,n}(y)|\frac{\sigma_{j,n}(y)}{\hat{\sigma}_{j,n}(y)}+B_{j,n}(y)+\frac{\sqrt{np_nh_n^{J-1}}\Delta^{(j)}(y)}{\hat{\sigma}_{j,n}(y)}\right|>\tilde{c}_{1,n}^S\right)\\
&\quad -o(1)\\
&\geq \Pr\left(|Z_{j^*,n}(y^*)|\frac{\sigma_{j^*,n}(y^*)}{\hat{\sigma}_{j^*,n}(y^*)}+B_{j^*,n}(y^*)+\frac{\sqrt{np_nh_n^{J-1}}\delta}{\hat{\sigma}_{j^*,n}(y^*)}>\tilde{c}_{1,n}^S\right)-o(1)\\
&\geq \Pr\left(\begin{array}{c}|Z_{j^*,n}(y^*)|\frac{\sigma_{j^*,n}(y^*)}{\hat{\sigma}_{j^*,n}(y^*)}+B_{j^*,n}(y^*)+\frac{\sqrt{np_nh_n^{J-1}}\delta}{\hat{\sigma}_{j^*,n}(y^*)}\\ >c_n(\alpha+\tau_n)+Cc_n(\gamma_n+\tau_n)-\epsilon_{6n}\end{array}\right)-o(1),
\end{aligned}
$$

where $(j^*,y^*)\in\arg\sup_{j\in\mathcal{J},y\in\mathcal{Y}_{n,L,c}}\Delta^{(j)}(y)$. The first inequality comes from Lemma A1(ii) and the Condition (iii) of Proposition 6. The last inequality comes from Condition SH2. By Condition (i) of Proposition 6, $\Delta^{(j^*)}(y^*)=\delta>0$. Then $\frac{\sqrt{np_nh_n^{J-1}}\delta}{\hat{\sigma}_{j^*,n}(y^*)}=O_p\left(\sqrt{np_nh_n^{J-1}}\right)$. Meanwhile $c_n(\alpha+\tau_n)\leq\frac{C_1\log n}{\alpha+\tau_n}$ and $c_n(\gamma_n+\tau_n)\leq\frac{C_1\log n}{\gamma_n+\tau_n}$ by Markov's inequality and Condition SH1(i). Proposition 6 follows if $np_nh_n^{J-1}\gamma_n/\log n\to\infty$. In particular, when $C=0$, Proposition 6 follows if $np_nh_n^{J-1}/\log n\to\infty$. **Q.E.D.**

# Appendix B

# Appendix for Chapter 3

## B.1 A Discussion of the Alternative Equilibrium Notion

To analyze the equilibrium characterizing equations given by the second notion of equilibrium. We reproduce (3.4) as follows. An equilibrium is a pair of cutoffs in the value of private information $(u_1^*(x), u_2^*(x))$ such that the following holds

$$\begin{aligned}
\pi_1(1,0,x_1) + \Delta_1(x_1)\Pr\left(U_{2i} \leq u_2^*(x)|X_i = x, U_{1i} = u_1^*(x)\right) &= u_1^*(x), \\
\pi_2(1,0,x_2) + \Delta_2(x_2)\Pr\left(U_{1i} \leq u_1^*(x)|X_i = x, U_{2i} = u_2^*(x)\right) &= u_2^*(x).
\end{aligned}$$

Moreover, let

$$\begin{aligned}
G_2(u_2^*, x, u_1^*) &= \Pr\left(U_{2i} \leq u_2^*|X_i = x, U_{1i} = u_1^*\right), \\
G_1(u_1^*, x, u_2^*) &= \Pr\left(U_{1i} \leq u_1^*|X_i = x, U_{2i} = u_2^*\right),
\end{aligned}$$

and

$$\psi(u^*, x) = \left[\begin{array}{c} u_1^* - \pi_1(1,0,x_1) - \Delta_1(x_1)G_2(u_2^*, x, u_1^*) \\ u_2^* - \pi_2(1,0,x_1) - \Delta_2(x_1)G_1(u_1^*, x, u_2^*) \end{array}\right].$$

The system of equilibrium characterizing equations is

$$\psi(u^*, x) = 0.$$

This system of equations corresponds to (3.7). Definitions 1, 2 and Assumption 3 can be modified by replacing $\varphi$ and vector $\sigma$ with $\psi$ and $u^*$ respectively. The observed conditional choice probabilities can be expressed as

$$Q_k\left(1|1,x\right) = \sum_{m \in \Upsilon(x)} \pi_m(x) \Pr\left(U_i \leq u_k^{*(m)}(x)|X_i = x, U_{-ki} \leq u_{-k}^{*(m)}(x)\right),$$

for $k = 1, 2$. This corresponds to (3.9) if the following Assumption 6 holds. Hence the results of this paper follows under this notion of equilibrium.

**Assumption 6.** For any $(a, b) \neq (a', b')$ satisfying

$$\begin{aligned} \Pr\left(U_{ki} \leq a | X_i = x, U_{-ki} \leq b\right) &\in (0,1), \text{ and} \\ \Pr\left(U_{ki} \leq a' | X_i = x, U_{-ki} \leq b'\right) &\in (0,1), \end{aligned}$$

for $k = 1, 2$, we have

$$\Pr\left(U_{ki} \leq a | X_i = x, U_{-ki} \leq b\right) \neq \Pr\left(U_{ki} \leq a' | X_i = x, U_{-ki} \leq b'\right).$$

Assumption 6 implies that a jump in the equilibrium cut-off $(u_1^*, u_2^*)$ must lead to a jump in the conditional belief $\Pr\left(U_{ki} \leq u_k^* | X_i = x, U_{-ki} \leq u_{-k}^*\right)$ for $k = 1, 2$, as long as the latent conditional belief takes a value strictly between 0 and 1.

## B.2 Mathematical Proofs

In this section, we present proofs for Propositions 1 to 3, Lemma 1 and Corollaries 1 and 2. The proof of Lemma 1 needs Lemma 2, which is also stated below.

**Proof of Proposition 1.**

We only need to consider Category (C1) because Sub-category (C2-1) can be viewed as a case in Category (C1). By (3.9), it suffices to show that the $M$ in Definition 1 equals 1. By (C1), there exists a unique function $q(x)$ such that for any $(\sigma, x) \in (0,1)^2 \times \text{ int}(\mathcal{X})$ satisfying $\varphi(\sigma, x) = 0$, $\sigma = q(x)$. We need to show that $q(x)$ is twice continuously differentiable on $\text{int}(\mathcal{X})$. Under Assumptions 1-3 and applying Implicit function theorem on an arbitrary point $(\sigma_0, x_0) \in (0,1)^2 \times \text{ int}(\mathcal{X})$, there are neighbourhoods $A$ of $\sigma_0$ and $B$ of $x_0$ on which (3.7) uniquely defines $\sigma$ as a function of $x$. That is, there is a function $\xi : B \to A$ such that $\varphi(\xi(x), x) = 0$ for all $x \in B$; for each $x \in B$, $\xi(x)$ is the unique solution to (3.7) lying in $A$, and $\xi(x)$ is twice continuously differentiable on $B$. Because $(\sigma_0, x_0)$ is arbitrary and $q(x)$ is the only solution to (3.7) by (C1), we must have $q(x) = \xi(x)$ for all $x \in \text{int}(\mathcal{X})$. Thus $q(x)$ is twice continuously differentiable on a neighbourhood of every $x \in \text{int}\mathcal{X}$. This immediately leads to the desired result.**Q.E.D.**

**Proof of Lemma 1.**

Let $S = \{x \in \mathcal{X} : (3.7)$ admits a unique solution in $\sigma\}$. Without loss of generality, assume that $S \cap B_1 \neq \varnothing$. Denote $S \cap B_1 = S_1$. By construction, $\Upsilon(S_1) = \{1\}$. By Lemma 2 (stated and proven below), there exists some $h \in \{1, ..., M\}$, such that $B_1 \cap B_h \neq \varnothing$. Denote $C_h = B_1 \cap B_h$. Without loss of generality, $\bar{S}_1 \cap \bar{C}_h \neq \varnothing$. (Otherwise, there will be another $k \in \{1, ..., M\}$ such that $B_1 \cap B_h \neq \varnothing$ and $\bar{S}_1 \cap \bar{C}_k \neq \varnothing$, where $C_k = B_1 \cap B_k$, all we need is $B_1 \backslash S_1$ is not empty, which holds by the conclusion of Lemma 2(ii).) Form a subset $D_h$ of $C_h$ such that $\bar{S}_1 \cap \bar{D}_h \neq \varnothing$ and $\Upsilon(D_h)$ is well defined. In this case, $\{1, h\} \subset \Upsilon(D_1)$. Therefore, $S_1$ and $D_h$ corresponds to $C_1$ and $C_2$ in Assumption 4H with $\Upsilon(S_1) = \{1\}$ while $\Upsilon(D_h) \supset \{1, h\}$. **Q.E.D.**

The proof of Lemma 1 needs the following Lemma 2.

**Lemma 2.** If the game is in Category (C2) or (C3), then the following holds. (i) The constant $M$ in Definition 1 is larger or equal than 2. (ii) Suppose that Assumption 5 holds, then for any $m \in \{1, ..., M\}$, there exists a $k \in \{1, ..., M\}$ and $k \neq m$, such that $B_m \cap B_k \neq \varnothing$.

**Proof of Lemma 2**

(i). When the game is in (C2-2) or (C3), there exists some $x_M \in \mathcal{X}$ such that $\varphi(\sigma_1, x_M) = \varphi(\sigma_2, x_M) = 0$ for some $\sigma_1 \neq \sigma_2$. If $M = 1$, there is a function $q : \mathcal{X} \to (0,1)^2$ such that $\sigma = q(x)$ for all $(\sigma, x) \in \mathcal{X} \times (0,1)^2$ satisfying $\varphi(\sigma, x) = 0$. But this means $\sigma_1 = q(x_M)$ and $\sigma_2 = q(x_M)$, which can not happen.

(ii). It suffices to show that for every $m \in \{1, ..., M\}$, there exists some $x \in B_m$ such that $\varphi(\sigma, x) = 0$ for more than one $\sigma \in (0,1)^2$ . Suppose it is not true, we have a $B_m$ such that $\varphi(\sigma, x) = 0$ has a unique solution in $\sigma$ for all $x \in B_m$. Consider an arbitrary $l \in \{j \neq m : \bar{B}_m \cap \bar{B}_j \neq \varnothing\}$ and let $x_0$ be an arbitrary point $x_0 \in \bar{B}_m \cap \bar{B}_l$. By assumption, $B_m \cap B_l = \varnothing$ for all $l \neq m$. Thus $x_0 \in \bar{B}_m \backslash B_m$. For an such $l$, we have two cases:

Case 1. $\lim_{x \in B_m, x \to x_0} q_m(x) - \lim_{x \in B_l, x \to x_0} q_l(x) = 0$ for all $x_0 \in \bar{B}_m \cap \bar{B}_l$.

Thus we can combine $B_m$ and $B_l$ to reduce the number $M$, this contradicts with the Definition 1 in which $M$ is the smallest number. In particular, let $B'_m = \mathrm{int}\big(B_m \cup B_l \cup (\bar{B}_m \cap \bar{B}_l)\big)$ and

$$q'_m(x) = \begin{cases} q_m(x), & x \in B_m, \\ q_l(x), & x \in B_l, \\ \lim_{x \in B_m, x \to x_0} q_m(x), & x_0 \in \bar{B}_m \cap \bar{B}_l. \end{cases}$$

Also, by Assumption 5, we know that $\bigtriangledown_\sigma(\sigma_0, x_0)$ is invertible, where $\sigma_0 = \lim_{x\in B_m, x\to x_0} q_m(x) = \lim_{x\in B_l, x\to x_0} q_l(x)$, then we can apply the Implicit Function Theorem to obtain that there are neighbourhoods $A$ of $\sigma_0$ and $B$ of $x_0$, and a unique function $\varsigma : B \to A$ such that $\varsigma(x_0) = \sigma_0$ and for each $x \in B$, $\varsigma(x_0)$ is the unique solution to (3.7) lying in $A$, $\varphi(\xi(x), x) = 0$ for $x \in B$, and $\varsigma(x)$ is twice continuously differentiable at $x_0$. Therefore, $q'_m(x)$ must coincide with $\varsigma(x)$ for $x \in B$, and $q'_m(x)$ is twice continuously differentiable.

Case 2. $|\lim_{x\in B_m, x\to x_0} q_m(x) - \lim_{x\in B_l, x\to x_0} q_l(x)| = \delta > 0$, for some $x_0 \in \bar{B}_m \cap \bar{B}_l$.

By continuity of $\varphi : (0,1)^2 \times \mathcal{X} \to \mathbf{R}^2$, we have

$$\begin{aligned} \varphi(\sigma_m, x_0) &= \varphi(\sigma_l, x_0) = 0, \\ \text{where } \sigma_m &= \lim_{x\to x_0, x\in B_m} q_m(x),\ \sigma_l = \lim_{x\to x_0, x\in B_l} q_l(x). \end{aligned}$$

There must be an $l$ satisfying the following: there is no $l'$ such that every component of $\sigma_{l'} - \sigma_m$ has the same sign as $\sigma_l - \sigma_m$ and $|\sigma_{l'} - \sigma_m| < |\sigma_l - \sigma_m|$. We focus on such an $l$. By Assumption 5(ii), the derivative $\bigtriangledown_\sigma(\sigma_l, x_0)$ is invertible. By the Implicit Function Theorem, for any neighbourhood $A$ of $\sigma_l$, there is a neighbourhood $B$ of $x_0$ and a function $\xi : B \to A$ such that $\xi(x_0) = \sigma_l$, $\varphi(\xi(x), x) = 0$ for $x \in B$, and $\xi(x)$ is continuous at $x_0$. Therefore, for any $\epsilon$, there exists an $\eta$ satisfying

$$|x - x_0| < \eta' \Longrightarrow |\xi(x) - \sigma_l| < \epsilon.$$

Then, for any $\epsilon$, all $x' \in B_m \cap B(x_0, \eta) \cap B$ satisfy $\varphi(\xi(x'), x') = 0$ and $|\xi(x') - \sigma_l| < \epsilon$. Thus $|\xi(x') - \sigma_m| > \delta - \epsilon$.

On the other hand, by continuity of $q_m$, for any $\epsilon'$, there exists an $\eta'$ satisfying

$$|x - x_0| < \eta' \Longrightarrow |q_m(x) - \sigma_m| < \epsilon'.$$

By choosing $\epsilon$ and $\epsilon'$ sufficiently small, we have $\xi(x') \neq q_m(x')$ for all $x' \in B_m \cap B(x_0, \eta \wedge \eta') \cap B$. This contradicts the assumption that $\varphi(\sigma, x) = 0$ has a unique solution in $\sigma$ for all $x \in B_m$. **Q.E.D.**

**Proof of Proposition 2.**

By Assumption 4H, there exists an $l \in \{1, ..., M\}$ such that for some $C_1, C_2 \subset B_l$, and $C_1 \cap C_2 = \varnothing$, $\bar{C}_1 \cap \bar{C}_2 \neq \varnothing$, we have $\Upsilon(C_1) \neq \Upsilon(C_2)$. Since $C_1, C_2 \subset B_l$, $l \in \Upsilon(C_1)$ and $l \in \Upsilon(C_2)$. Then consider an arbitrary $x_d \in \bar{C}_1 \cap \bar{C}_2$, by (3.9),

$$\begin{aligned}
\lim_{x\to x_d, x\in C_1} Q\left(1|1,x\right) &= \sum_{h\in\Upsilon(C_1)} \lim_{x\to x_d, x\in C_1} \pi_h(x) \lim_{x\to x_d, x\in C_1} \sigma^{(h)}(1|1,x), \\
\lim_{x\to x_d, x\in C_2} Q\left(1|1,x\right) &= \sum_{j\in\Upsilon(C_2)} \lim_{x\to x_d, x\in C_2} \pi_j(x) \lim_{x\to x_d, x\in C_2} \sigma^{(j)}(1|1,x).
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \lim_{x\to x_d, x\in C_1} Q\left(1|1,x\right) - \lim_{x\to x_d, x\in C_2} Q\left(1|1,x\right) \\
= & \sum_{l\in\Upsilon(C_1)\cap\Upsilon(C_2)} \left( \lim_{x\to x_d, x\in C_1} \pi_l(x) - \lim_{x\to x_d, x\in C_2} \pi_l(x) \right) \sigma_k^{(l)}(1|1,x_d) \\
& + \sum_{h\in\Upsilon(C_1)\backslash\Upsilon(C_2)} \lim_{x\to x_d, x\in C_1} \pi_h(x) \lim_{x\to x_d, x\in C_1} \sigma_k^{(h)}(1|1,x) \\
& - \sum_{j\in\Upsilon(C_2)\backslash\Upsilon(C_1)} \lim_{x\to x_d, x\in C_2} \pi_j(x) \lim_{x\to x_d, x\in C_2} \sigma_k^{(j)}(1|1,x), \qquad \text{(B.1)}
\end{aligned}$$

where

$$\begin{aligned}
\sigma^{(l)}(1|1,x_d) &\neq \lim_{x\to x_d, x\in C_1} \sigma^{(h)}(1|1,x), \\
\sigma^{(l)}(1|1,x_d) &\neq \lim_{x\to x_d, x\in C_2} \sigma^{(j)}(1|1,x),
\end{aligned}$$

for any $l \in \Upsilon(C_1) \cap \Upsilon(C_2), h \in \Upsilon(C_1)\backslash\Upsilon(C_2)$, $j \in \Upsilon(C_2)\backslash\Upsilon(C_1)$, and for any $x_d \in \bar{C}_1 \cap \bar{C}_2$. Suppose to the contrary, for example, $\sigma^{(l)}(1|1,x_d) = \lim_{x\to x_d, x\in C_2} \sigma^{(h)}(1|1,x)$ for some $x_d \in \bar{C}_1 \cap \bar{C}_2$, this means $\lim_{x\to x_d, x\in C_1} q_l(x) = \lim_{x\to x_d, x\in C_2} q_h(x)$ for some $x_d \in \bar{C}_1 \cap \bar{C}_2$. Note that $x_d \in B_l$ and $x_d \in \bar{B}_h$, thus we have $q_l(x') = q_h(x')$ for some $x' \in B(x_d, r) \cap B_l \cap B_h$, where $B(x_d, r)$ is a ball centred in $x_d$ and with radius $r$. $B(x_d, r) \cap B_l \cap B_h \neq \varnothing$ because $x_d \in \bar{C}_1 \cap \bar{C}_2, C_1, C_2 \subset B_l$ and $C_2 \subset B_h$. This contradicts with Remark 1 below Definition 1.

Furthermore, observe that the set

$$\{(\lim_{x\in C_1, x\to x_d} \pi(x), \lim_{x\in C_2, x\to x_d} \pi(x)) : \pi \in \mathcal{C}_{E_1}(\Upsilon)\}$$

is

$$\{(\lim_{x\in C_1, x\to x_d} \pi(x), \lim_{x\in C_2, x\to x_d} \pi(x)) : \pi \in \mathcal{C}(\Upsilon) \text{ and (B.1)} = 0\},$$

which has zero measure in

$$\{(\lim_{x\in C_1, x\to x_d}\pi(x), \lim_{x\in C_2, x\to x_d}\pi(x)) : \pi\in\mathcal{C}(\Upsilon)\},$$

for every $x_d \in \bar{C}_1 \cap \bar{C}_2$ and for all $C_1, C_2, l$ satisfying Assumption 4H. The desired result follows. **Q.E.D.**

**Proof of Corollary 1.**

Consider a subset $D$ satisfying the property in Definition 3 and let $x_d$ be an arbitrary point in $\bar{D}_1 \cap \bar{D}_2$. Since $\Upsilon(D)$ is defined, i.e. the equilibrium indices set is the same within $D$, we have $x_d \in B_h$ for all $h \in \Upsilon(D)$. By (3.9),

$$\begin{aligned}
\lim_{x\to x_d, x\in D_1} Q(1|1,x) &= \sum_{h\in\Upsilon(D)} \lim_{x\to x_d, x\in D_1}\pi_h(x) \lim_{x\to x_d, x\in D_1}\sigma^{(h)}(1|1,x),\\
\lim_{x\to x_d, x\in D_2} Q(1|1,x) &= \sum_{h\in\Upsilon(D)} \lim_{x\to x_d, x\in D_2}\pi_h(x) \lim_{x\to x_d, x\in D_2}\sigma^{(h)}(1|1,x),
\end{aligned}$$

where $\lim_{x\to x_d, x\in D_1}\sigma^{(h)}(1|1,x) = \lim_{x\to x_d, x\in D_2}\sigma^{(h)}(1|1,x) = \sigma^{(h)}(1|1,x_d)$ for all $h\in\Upsilon(D)$. Thus

$$\begin{aligned}
&\left|\lim_{x\to x_d, x\in D_1} Q(1|1,x) - \lim_{x\to x_d, x\in C_2} Q(1|1,x)\right|\\
&= \sum_{h\in\Upsilon(D)}\left(\lim_{x\to x_d, x\in D_1}\pi_h(x) - \lim_{x\to x_d, x\in D_2}\pi_h(x)\right)\sigma^{(h)}(1|1,x_d)\\
&= \sum_{h\in\Upsilon_1(D)}\left(\lim_{x\to x_d, x\in D_1}\pi_h(x) - \lim_{x\to x_d, x\in D_2}\pi_h(x)\right)\sigma^{(h)}(1|1,x_d), \quad \text{(B.2)}
\end{aligned}$$

where

$$\Upsilon_1(D) = \left\{\ h\in\Upsilon(D) : \left|\lim_{x\to x_d, x\in D_1}\pi_h(x) - \lim_{x'\to x_d, x'\in D_2}\pi_h(x')\right| > 0\ \right\}.$$

Let $\mathcal{C}_F(\Upsilon)$ be the subset of $\mathcal{C}(\Upsilon)$ that satisfies the property in Definition 3. Furthermore, observe that the set

$$\{(\lim_{x\in D_1, x\to x_d}\pi(x), \lim_{x\in D_2, x\to x_d}\pi(x)) : \pi\in\mathcal{C}_{E_2}(\Upsilon)\}$$

is

$$\{(\lim_{x\in D_1, x\to x_d}\pi(x), \lim_{x\in D_2, x\to x_d}\pi(x)) : \pi\in\mathcal{C}_F(\Upsilon) \text{ and (B.2)} = 0\},$$

which has a zero measure in

$$\{(\lim_{x\in D_1, x\to x_d} \pi(x), \lim_{x\in D_2, x\to x_d} \pi(x)) : \pi \in \mathcal{C}_F(\Upsilon)\},$$

for every $x_d \in \bar{D}_1 \cap \bar{D}_2$ in Definition 3. The desired result follows. **Q.E.D.**

**Proof of Corollary 2.**

When the game is in sub-category (C2-2), there exists two subsets of $\mathcal{X}$, $C_1$ and $C_2$, $\bar{C}_1 \cap \bar{C}_2 \neq \varnothing$, and $m_1^* \neq m_2^*$ such that $\pi_{m_1^*}(x) = 1$ for $x \in C_1$ and $\pi_{m_2^*}(x) = 1$ for $x \in C_2$. For any $x_d \in \bar{C}_1 \cap \bar{C}_2$, there are two cases:

Case 1. There is an $r > 0$ such that $\Upsilon(B(x_d, r))$ is defined, where $B(x_d, r)$ is an open ball centred in $x_d$ and with radius $r$. As the equilibrium indices set does not change on $B(x_d, r)$, the ball $B(x_d, r) \subset B_m$ if $m \in \Upsilon(B(x_d, r))$. Therefore,

$$\begin{aligned} & \lim_{x\to x_d, x\in C_1\cap\Upsilon(B(x_d,r))} Q\,(1|1,x) \\ = & \lim_{x\to x_d, x\in C_1\cap\Upsilon(B(x_d,r))} \sigma^{(m_1^*)}(1|1,x) = \sigma^{(m_1^*)}(1|1,x_d), \end{aligned}$$

where the second equality comes from $B(x_d, r) \subset B_m$ if $m \in \Upsilon(B(x_d, r))$. Similarly,

$$\begin{aligned} & \lim_{x\to x_d, x\in C_2\cap\Upsilon(B(x_d,r))} Q\,(1|1,x) \\ = & \lim_{x\to x_d, x\in C_2\cap\Upsilon(B(x_d,r))} \sigma^{(m_2^*)}(1|1,x) = \sigma^{(m_2^*)}(1|1,x_d). \end{aligned}$$

Since $\sigma^{(m_1^*)}(1|1,x_d) \neq \sigma^{(m_2^*)}(1|1,x_d)$ (Otherwise, Remark 1 of Definition 1 is violated), $Q\,(1|1,x)$ has a jump at $x_d$.

Case 2. There is no such $r > 0$ that $\Upsilon(B(x_d, r))$ is defined. That is, the equilibrium indices set changes at $x_d$. Then there exist disjoint subsets $D_1, D_2$ such that $x_d \in \bar{D}_1 \cap \bar{D}_2$, $D_1 \subset C_1$, $D_2 \subset C_2$, $\Upsilon(D_1)$ and $\Upsilon(D_2)$ are defined, and $\Upsilon(D_1) \neq \Upsilon(D_2)$. We have

$$\begin{aligned} \lim_{x\to x_d, x\in D_1} Q\,(1|1,x) &= \lim_{x\to x_d, x\in D_1} \sigma^{(m_1^*)}(1|1,x), \\ \lim_{x\to x_d, x\in D_2} Q\,(1|1,x) &= \lim_{x\to x_d, x\in D_2} \sigma^{(m_2^*)}(1|1,x). \end{aligned}$$

Since the choice of $x_d \in \bar{C}_1 \cap \bar{C}_2$ is arbitrary, we must have

$$\lim_{x\to x_d, x\in D_1} \sigma_k^{(m_1^*)}(1|1,x) \neq \lim_{x\to x_d, x\in D_2} \sigma_k^{(m_2^*)}(1|1,x),$$

for some $x_d$. Otherwise $B_{m_1^*}$ and $B_{m_2^*}$ can be combined together so that $M$ is not the smallest constant in Definition 1. Therefore, $Q_k\left(1|1,x\right)$ has a jump at $x_d$. **Q.E.D.**

**Proof of Proposition 3.**

(i) By (3.15), we have

$$Q\left(1|1,x\right)=\sum_{t=1}^{\bar{t}}\lambda_t Q^t\left(1|1,x\right).$$

By Proposition 1, $Q^t\left(1|1,x\right)$ is twice continuously differentiable in $x$ for all $t\in\{1,...,\bar{t}\}$. Thus $Q_k\left(1|1,x\right)$ is twice continuously differentiable in $x$.

(ii) For an $x_d$ satisfying the condition COND in Proposition 3 (ii), by (3.15)

$$Q\left(1|1,x\right)=\sum_{t_1\in\{t':x_d\in\mathcal{X}_D^{t'}\}}\lambda_{t_1}Q^{t_1}\left(1|1,x\right)+\sum_{t_2\in\{t':x_d\notin\mathcal{X}_D^{t'}\}}\lambda_{t_2}Q^{t_2}\left(1|1,x\right). \quad \text{(B.3)}$$

By construction, $Q^{t_2}\left(1|1,x\right)$ does not have a jump at $x=x_d$, for all $t_2\in\{t':x_d\notin\mathcal{X}_D^{t'}\}$. As a result, the second summation on the right hand side of (B.3) does not have a jump at $x=x_d$. By COND, the first summation on the right hand side of (B.3) has a jump at $x=x_d$. Putting the two summations together, $Q\left(1|1,x\right)$ must have a jump at $x=x_d$.

(iii) can be shown in the same way as (ii). **Q.E.D.**