GOVERNANCE AND THE EFFICIENCY OF ECONOMIC SYSTEMS
GESY

Discussion Paper No. 432

# A Semiparametric Test of Agent's Information Sets for Games of Incomplete Information

Salvador Navarro *
Yuya Takahashi **

* Department of Economics, Social Science Centre, Canada
** University of Mannheim

March 2012

Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

Sonderforschungsbereich/Transregio 15 · www.sfbtr15.de
Universität Mannheim · Freie Universität Berlin · Humboldt-Universität zu Berlin · Ludwig-Maximilians-Universität München
Rheinische Friedrich-Wilhelms-Universität Bonn · Zentrum für Europäische Wirtschaftsforschung Mannheim

Speaker: Prof. Dr. Klaus M. Schmidt · Department of Economics · University of Munich · D-80539 Munich,
Phone: +49(89)2180 2250 · Fax: +49(89)2180 3510

# A Semiparametric Test of Agent's Information Sets for Games of Incomplete Information


Salvador Navarro and Yuya Takahashi*

University of Western Ontario and University of Mannheim


March 15, 2012


**Abstract**

We propose semiparametric tests of misspecification of agent's information for games of incomplete information. The tests use the intuition that the opponent's choices should not predict a player's choice conditional on the proposed information available to the player. The tests are designed to check against some commonly used null hypotheses (Bajari et al. (2010), Aradillas-Lopez (2010)). We show that our tests have power to discriminate between common alternatives even in small samples. We apply our tests to data on entry in the US airline industry. Both the assumptions of independent and correlated private shocks are not supported by the data.



*Corresponding address: Department of Economics, Social Science Centre, Room 4037, London, Ontario, N6A 5C2, Canada. Tel: +1 519 661 2111 ext 81586, Fax: +1 519 661 3666.

*E-mail address:* s.navarro.lozano@gmail.com (S. Navarro), ytakahas@mail.uni-mannheim.de (Y. Takahashi).



We thank Jaap Abbring, Victor Aguirregabiria, Andres Aradillas-Lopez, Pat Bajari, Jean-Francois Houde, Aureo de Paula, Jack Porter and participants at the 2009 IIOC meetings, the eight invitational choice symposium and the 2011 AEA meetings. First draft: March, 2009. This draft: March, 2012.

# 1 Introduction

There is a growing literature on the estimation of games with incomplete information (e.g., Brock and Durlauf (2001), Seim (2006), Sweeting (2009), Bajari, Hong, Krainer, and Nekipelov (2010) and Aradillas-Lopez (2010) for static games and Aguirregabiria and Mira (2007), Pesendorfer and Schmidt-Dengler (2008), Collard-Wexler (2010), Sweeting (2011), and Ryan (2011) in the literature on the estimation of dynamic games). Because incomplete information can take many forms, it is common for the analyst to simply choose some information structure and analyze the game under this maintained assumption. A convenient and common assumption is that the payoff shocks that are unobservable to the econometrician are private information from the player's perspective. This assumption effectively imposes the restriction that each player participating in the game has access to the same information about its competitors as the outside observer analyzing the situation (i.e. the econometrician). In this case, the equilibrium choice probabilities that the analyst can recover from the data as a function of observable covariates coincide with the player's equilibrium beliefs. Hence, this assumption effectively simplifies the estimation problem of strategic interactions to one of a single agent random utility model.

While convenient, there is no a priori reason to believe that a player and the econometrician have the same amount of information about the player's competitors. In particular, it is likely that the payoff shocks unobserved to the econometrician are at least partially observed by the agents participating in the game. Partially observed players' shocks invalidate the strategy of estimating equilibrium beliefs directly from the conditional choice probabilities and generate dependence among players' choices. This misspecification of information on the part of the econometrician will lead to biased estimates and mistaken inference.[1]

As a first step in dealing with this potential problem, this paper proposes two simple semiparametric specification tests of the hypothesis that payoff shocks unobserved to the econometrician are entirely private information. Since one of the main advantages of assuming that the player's and econometrician's information (about competitors) coincide is the simplicity of the resulting estimators, we propose a test that is equally simple. This first test assumes that realizations of payoff shocks are *iid* among players and tests against the hypothesis that payoff shocks are entirely private information. The logic behind this test is simple: under the *iid* assumption, if players partially observe their opponents' shocks but the econometrician does not, then the players' observed equilibrium choices will not be independent of each other, even after controlling for the observable (to the econometrician) covariates. Thus, the test checks for dependence among players' choices after controlling for observable covariates. If dependence is detected the null hypothesis that players use the same information as the econometrician when inferring their competitors' decisions is rejected.[2]

[1] See Cunha et al. (2005) for a similar point in the context of a lifecycle model with no strategic interactions.

[2] Although not exactly the same, the question we ask is isomorphic to the one in Heckman and Navarro (2004) where

The second test we propose allows for the possibility that realizations of payoff shocks among players are exogenously correlated as in Aradillas-Lopez (2010). Under this correlation structure, our procedure tests the null hypothesis that payoff shocks are entirely private information. If shocks are correlated and players know the joint distribution of the shocks, a player's realization of his own shock will help him when forming expectations about his opponents' shocks (i.e. a signal extraction problem). In this case, dependence (conditional on observable covariates) can come from partial observability and/or from exogenous correlation of shocks. Therefore, we need to control for the latter factor (exogenous correlation) to test whether there is partial observability. Under the null hypothesis of correlated shocks but no partial observability, other players' choices net of the effect of observables, i.e. their unobservable (to the econometrician) shocks, should be independent of the current player's choice. Since our test now relies on including unobservable shocks when estimating probabilities, our proposed method jointly estimates auxiliary testing parameters and the joint distribution of all players' unobservables.

This paper is related to Grieco (2010). In a similar spirit as ours, he proposes a flexible information structure that nests as a special case the private information assumption that many papers place. He proves that this assumption is testable based on independence of private payoff shocks and exclusion restrictions. Unlike Grieco (2010), our focus is on testing procedures. Thus, our test is easy to implement and requires none of these assumptions. In particular, our second test relaxes the independence of private shocks, which is a significant step towards a general framework. Our work also relates to de Paula and Tang (2011), who use the same intuition as our test in order to test the existence of multiple equilibria. Their logic is that, with non-deterministic equilibrium selection rules, multiple equilibria break the conditional independence assumption. As opposed to them, we assume a deterministic (conditional on observables) equilibrium selection rule, hence we interpret the failure of conditional independence as a rejection of the null of entirely private information. Sweeting (2009) performs a test to examine whether there is any time-series correlation in players' actions in the same market, which is evidence against private information. Since his test is specific to his application in that it requires time-series variations and many players in the same market, our first test can be regarded as a more general and easy-to-implement version of the test in Sweeting (2009).

The rest of the paper proceeds as follows. In section 2 we lay down a simple two player game with incomplete information in which each player makes a binary decision. We then characterize the 3 different sets of assumptions about information we test for in section 3. In section 4 we develop the tests and show their power properties via Monte Carlo simulation. We apply our test to data on entry in the US airline industry in section 5. Section 6 concludes.

---

they characterize the informational requirements of methods that control for selection only based on variables observed by the econometrician.

## 2 A Simple Two Player Game with Binary Actions

Consider a game of incomplete information where two players, $i$ and $j$ have to choose one of two possible actions.[3] Let $S_i$ denote all the random variables affecting player $i$'s payoff regardless of whether they are observed by both players and/or the econometrician. A simple example would be a two firm entry game where $S_i$ would denote the variables determining firm $i$'s profit.[4] Divide $S_i = (X_i, \varepsilon_i)$ where $X_i$ is the set of variables observable to both players and the econometrician and $\varepsilon_i$ is vector of random variables unobserved to the econometrician but observed by player $i$.[5] The extent to which $\varepsilon_i$ is observed by player $j$ is what we wish to determine.

Let $a_i$ denote player $i$'s action, and let the set of actions be denoted by $A_i = \{0, 1\}$. For simplicity, denote $a_{-i} = a_j$ and $A_{-i} = A_j$. Player $i$'s payoff depends also on his own choice and his rival's choice. Formally we write the payoff as

$$u_i(a_i, a_{-i}, S_i) = U_i(a_i, a_{-i}, X_i) - \varepsilon_i(a_i), \tag{1}$$

where we allow $\varepsilon_i$ to (potentially) depend on the action taken by player $i$. We assume that the payoff "shock" ($\varepsilon_i$) is independent of all the observable covariates.[6] We further assume that both players draw the random shock $\varepsilon$ from the common and known distribution $G_\varepsilon$, which is absolutely continuous with unbounded support and density $g > 0$ everywhere.

$\varepsilon_i$ and $\varepsilon_j$ are both unobserved to the econometrician, but we allow for the possibility that part of $\varepsilon_i$ is observed by player $j$ and part of $\varepsilon_j$ is observed by player $i$. We further allow for the possibility that what player $i$ observes about player $j$ is different from what player $j$ observes about $i$ so there can be informational asymmetries between players, i.e. the potential partial observability is not necessarily due to a common shock.

In order to fix ideas we further specialize the framework and work with a simple example. Consider a simple static entry model where 2 players simultaneously choose between entering or not. Entry of player $j$ affects (arguably reduces) player $i$'s profit. Without loss of generality, we normalize the profit of not entering to zero for both players. Specifically, we assume that profits are given by

$$\Pi_i = \begin{cases} h_i(X_i) + \alpha_i y_j - \varepsilon_i & \text{if } y_i = 1 \\ 0 & \text{if } y_i = 0 \end{cases}, \tag{2}$$

[3]Extending the game (and the tests) to an $n$-player case and/or $m$-alternative case is straightforward at the cost of considerable notational burden. Neither our tests nor any of the points we make depend on the simple setup we use in this section.

[4]See Bresnahan and Reiss (1991), Berry (1992), Mazzeo (2002) and Seim (2006) for examples.

[5]We can also make $X_i$ unobservable to the econometrician and introduce an observable signal for $X_i$ instead as in Aradillas-Lopez (2010).

[6]In Section 3.4 we discuss how we can relax this assumption.

where $y_j = 1$ if player $j$ enters the market and $y_j = 0$ otherwise. If we let $\Omega_i$ denote player $i$'s information set (i.e. its state variables at time $t$) and let $\pi_j \equiv E(y_j = 1|\Omega_i)$, the optimal choices are then given by

$$y_i = \mathbb{1}\{h_i(X_i) + \alpha_i \pi_j - \varepsilon_i \geq 0\}, \tag{3}$$

where $\mathbb{1}\{a\}$ is an indicator function that equals one if $a$ is true, and zero otherwise.

## 2.1 Alternative Information Structures

We consider three alternative information structures (i.e. specifications for $\Omega_i$) for a game of the kind described above. The first one is the independent private shocks (IPS) specification, in which it is assumed that $\varepsilon_i$ and $\varepsilon_j$ are *iid* and entirely each player's private information. Bajari et al. (2010) assume this shock structure to estimate a discrete game of incomplete information. The second specification we consider is the correlated private shocks (CPS) specification, in which it is assumed that, while $\varepsilon_i$ and $\varepsilon_j$ are private information, they may be correlated with each other. Because players are assumed to know the joint distribution of $\varepsilon_i$ and $\varepsilon_j$, each player conditions on the realization of his own $\varepsilon$ when forming expectations about his opponent's entry probability. Aradillas-Lopez (2010) provides a framework of estimating a discrete game of incomplete information under this general shock structure. The third information structure we propose in this paper assumes that $\varepsilon_i$ and $\varepsilon_j$ are independent but we allow for the possibility that player $i$ partially observes $\varepsilon_j$ and that player $j$ partially observes $\varepsilon_i$.

### 2.1.1 Independent Private Shocks (IPS)

In this case the information set for player $i$ is given by $\Omega_i = (X_i, X_j, \varepsilon_i)$. A Bayesian-Nash equilibrium is given by a set of optimal strategies and beliefs consistent with these strategies. That is, a Bayesian-Nash equilibrium of this game is given by

$$y_1 = \mathbb{1}\{h_1(X_1) + \alpha_1 \pi_2^* - \varepsilon_1 \geq 0\} \tag{4}$$

$$y_2 = \mathbb{1}\{h_2(X_2) + \alpha_2 \pi_1^* - \varepsilon_2 \geq 0\}, \tag{5}$$

where $(\pi_1^*, \pi_2^*)$ is a fixed point of $\varphi = (\varphi_1, \varphi_2) = \mathbf{0}$ with

$$\varphi_1(\pi_1, \pi_2) = \pi_1 - G_{\varepsilon_1}(h_1(X_1) + \alpha_1 \pi_2) \tag{6}$$

$$\varphi_2(\pi_1, \pi_2) = \pi_2 - G_{\varepsilon_2}(h_2(X_2) + \alpha_2 \pi_1). \tag{7}$$

Equations (6) and (7) imply that both $\pi_1^*$ and $\pi_2^*$ are functions of only $\mathbf{X} = (X_1, X_2)$.[7] We explicitly denote this dependence by writing $\pi_1^* = \pi_1(\mathbf{X})$ and $\pi_2^* = \pi_2(\mathbf{X})$. The fact that the equilibrium probabilities are a function only of the observables $\mathbf{X}$ is the key result that we use when designing our test of whether an agent knows some (or all) of his opponents' $\varepsilon$.

### 2.1.2 Correlated Private Shocks (CPS)

Let $G_{\varepsilon_1,\varepsilon_2}(\cdot,\cdot)$ be the joint distribution of $(\varepsilon_1, \varepsilon_2)$ and let $g_{\varepsilon_1|\varepsilon_2}(\varepsilon_1|\varepsilon_2)$ denote the density of $\varepsilon_1$ conditional on $\varepsilon_2$. As shown in Aradillas-Lopez (2010), since now the realization of the privately observed shock $\varepsilon_1$ contains information about the realized $\varepsilon_2$, the equilibrium beliefs will be functions of shock realizations. That is, a Bayesian-Nash equilibrium of this game is given by

$$\begin{aligned} y_1 &= \mathbb{1}\{h_1(X_1) + \alpha_1 \pi_2^* - \varepsilon_1 \geq 0\} && (8) \\ y_2 &= \mathbb{1}\{h_2(X_2) + \alpha_2 \pi_1^* - \varepsilon_2 \geq 0\}, && (9) \end{aligned}$$

where $(\pi_1^*, \pi_2^*)$ is a solution to the following system of functional equations:

$$\begin{aligned} \pi_1^*(\mathbf{X}, \varepsilon_2) &= \int \mathbb{1}\{h_1(X_1) + \alpha_1 \pi_2^*(\mathbf{X}, \varepsilon_1) - \varepsilon_1 \geq 0\} g_{\varepsilon_1|\varepsilon_2}(\varepsilon_1|\varepsilon_2) \, d\varepsilon_1 && (10) \\ \pi_2^*(\mathbf{X}, \varepsilon_1) &= \int \mathbb{1}\{h_2(X_2) + \alpha_2 \pi_1^*(\mathbf{X}, \varepsilon_2) - \varepsilon_2 \geq 0\} g_{\varepsilon_2|\varepsilon_1}(\varepsilon_2|\varepsilon_1) \, d\varepsilon_2. && (11) \end{aligned}$$

Note that, even after controlling for the observables $\mathbf{X}$, player $i$'s beliefs about player $j$'s probability of entry $(\pi_j^*)$ depend on player $i$'s shock but not on $\varepsilon_j$. The fact that beliefs will not depend on $\varepsilon_j$ is the key to the second test we develop below.

### 2.1.3 Partially Observable Shocks (POS)

The final information specification we consider assumes that $\varepsilon_i$ is potentially *partially observable* by the opposing player. That is, we allow for the possibility that part (or all) of $\varepsilon_i$ is observed to $i$'s opponent. For simplicity, we assume that the shock can be decomposed in an additive form:[8]

$$\varepsilon_i = \varepsilon_i^o + \varepsilon_i^u, \qquad (12)$$

[7] In case of multiple equilibria $\pi_1^*$ and $\pi_2^*$ are correspondences. We come back to this issue in section 3.3.

[8] We assume additivity for simplicity in order to generate data in our simulations. Clearly any function

$$\varepsilon_i = f_i(\varepsilon_i^o, \varepsilon_i^u)$$

will have the same implications.

where $\varepsilon_i^o$ is observed to $i$'s opponent, and $\varepsilon_i^u$ is observed only to $i$. Neither $\varepsilon_i^o, \varepsilon_i^u$ nor $\varepsilon_i$ are observed by the econometrician. In terms of the notation introduced before, $i$'s information set would be given by $\Omega_i = \left(X_i, X_j, \varepsilon_i, \varepsilon_j^o\right)$. Assume that $\varepsilon_1^o$, $\varepsilon_1^u$, $\varepsilon_2^o$, and $\varepsilon_2^u$ are all mutually independent.

Under these assumptions, the equilibrium beliefs are functions of shock realizations too. A Bayesian-Nash equilibrium of this game is given by

$$y_1 = \mathbb{1}\{h_1(X_1) + \alpha_1 \pi_2^* - \varepsilon_1^o - \varepsilon_1^u \geq 0\} \tag{13}$$

$$y_2 = \mathbb{1}\{h_2(X_2) + \alpha_2 \pi_1^* - \varepsilon_2^o - \varepsilon_2^u \geq 0\}, \tag{14}$$

where $(\pi_1^*, \pi_2^*)$ is a solution to the following system of equations:

$$\pi_1^*(\mathbf{X}, \varepsilon_1^o, \varepsilon_2^o) = \int \mathbb{1}\{h_1(X_1) + \alpha_1 \pi_2^*(\mathbf{X}, \varepsilon_1^o, \varepsilon_2^o) - \varepsilon_1^o - \varepsilon_1^u \geq 0\}\, g_{\varepsilon_1^u}(\varepsilon_1^u)\, d\varepsilon_1^u \tag{15}$$

$$\pi_2^*(\mathbf{X}, \varepsilon_1^o, \varepsilon_2^o) = \int \mathbb{1}\{h_2(X_2) + \alpha_2 \pi_1^*(\mathbf{X}, \varepsilon_1^o, \varepsilon_2^o) - \varepsilon_2^o - \varepsilon_2^u \geq 0\}\, g_{\varepsilon_2^u}(\varepsilon_2^u)\, d\varepsilon_2^u. \tag{16}$$

The key thing to notice is that, under partial observability, player $i$'s equilibrium beliefs will depend on the realization of his opponent's shock, even after controlling for observables and for his own shock.

## 3 Semiparametric Specification Tests

In this section we introduce the specification tests that will allow us to distinguish between the 3 models just presented. Because the key aspect that we wish to test for is the specification of $\Omega$ and not to recover the structural model, the tests we develop are semiparametric in their specification of the payoff functions. That is, while in our discussion of the models we assumed additive separability between the direct payoff $(h_i)$, the strategic interaction term $(\alpha_i \Pr(y_j|\Omega_i))$ and the shocks, the test are general enough to allow for models specified under weaker nonseparable payoffs.[9] We impose the following assumptions:

***A-1*** *(Data) Let $F_{Y_1,Y_2}(y_1, y_2|\mathbf{X})$ be the joint distribution of $(y_1, y_2)$ conditional on $\mathbf{X}$. The econometrician has access to a large number of repetitions of games so that $F_{Y_1,Y_2}(y_1, y_2|\mathbf{X})$ can be treated as known.*

***A-2*** *(DGP) Data is generated from one of the three models described in the previous section. The econometrician doesn't know the true model.*

[9]To be specific, we apply our tests in the context of the information structures described above (see assumption ***A-2***). However, the tests we propose can apply more generally (even for certain classes of dynamic games). The only requirement is that the policy functions that arise as an equilibrium of the game are functions of the specified (a priori) information available to each agent. With this in hand, we can simply follow the same strategy of adding the "left-out" information and testing for its predictive power.

***A-3*** *(Multiple equilibria) Multiple equilibria are allowed but we assume the existence of a deterministic equilibrium selection rule. The rule assigns an equilibrium based on public information. The econometrician does not need to know the rule, but players do.*

de Paula and Tang (2011) relax A-3 and account for cases in which the equilibrium selection rule is not deterministic. Aradillas-Lopez and Gandhi (2011) do not specify the nature of equilibrium selection when considering inference of parameters in ordered response games with incomplete information. Both papers, however, maintain the assumption of independent private shocks. See section 3.3 for a discussion on the issue of multiple equilibria and possible alternative assumptions to ***A-3***. In addition, we allow for the possibility that $X_1 = X_2$, which means that we do not rely on exclusion restrictions.

## 3.1 Null Hypothesis: Independent Private Shocks

We first consider the testable implications of assuming the IPS specification. In this case, both $\pi_1^*$ and $\pi_2^*$ are just functions of $\mathbf{X}$ and hence (4) can be written as

$$\begin{aligned} y_1 &= \mathbb{1}\{h_1(X_1) + \alpha_1 \pi_2^*(\mathbf{X}) - \varepsilon_1 \geq 0\} \\ &= \mathbb{1}\{\mu_1(\mathbf{X}) - \varepsilon_1 \geq 0\}, \end{aligned} \tag{17}$$

for some function $\mu$.[10] The null and alternative hypotheses are

$H_0$ : shocks are *iid* and private information

$H_1$ : shocks are correlated or partially observed.

To make the test operational we take advantage of the fact that, under $H_0$, $y_1$ and $y_2$ are assumed to be independent random variables once we control for $\mathbf{X}$. Therefore, we consider the following testing equation[11]:

$$y_1 = \mathbb{1}\{\mu_1(\mathbf{X}) + \delta_1 y_2 - \varepsilon_1 \geq 0\}. \tag{18}$$

where $\delta_1$ is an auxiliary parameter to be used for testing purposes. The key idea behind the test is that,

[10] The second line makes it clear that we don't strictly require (4) to be the data generating process. Our test, will apply to any model with the same information structure that generates the second line of (17).

[11] Bajari et al. (2010) also consider a model with market fixed effects. However, they assume that the market level unobservable is just a function of observable covariates. Hence, for market $m$, (17) is rewritten as

$$\begin{aligned} y_{1m} &= \mathbb{1}\{h_1(X_{1m}) + \alpha_1 \pi_2^*(\mathbf{X}_m) + \eta(\mathbf{X}_m) - \varepsilon_{1m} \geq 0\} \\ &= \mathbb{1}\{\widetilde{\mu}_1(\mathbf{X}_m) - \varepsilon_{1m} \geq 0\}, \end{aligned}$$

implying that our testing procedure (18) is still valid even in this case.

under the null hypothesis, $\delta_1 = 0$.[12] So we consider the following hypothesis instead:

$$H_0' \quad : \quad \delta_1 = 0 \tag{19}$$

$$H_1' \quad : \quad \delta_1 \neq 0, \tag{20}$$

where rejection of $H_0'$ implies the rejection of $H_0$.

Notice that the test we propose can be easily implemented as a $t$-test of significance of the auxiliary parameter $\delta_1$. One can also choose to include a more general auxiliary function of $y_2$.[13] As we show below in our simulations, the test performs as expected under the null (i.e. we cannot reject $\delta_1 = 0$). More important, as we also show, the power of the test (i.e. its ability to reject the null when it is false) is remarkably good both against the CPS and the POS alternatives.

### 3.2 Null Hypothesis: Correlated Private Shocks

When the true data generating process is given by the CPS model, both $\pi_1^*$ and $\pi_2^*$ are functions not only of $\mathbf{X}$ but of $\varepsilon_2$ and $\varepsilon_1$, respectively. Hence, $y_1$ and $y_2$ may be correlated even after controlling for $\mathbf{X}$. However, once we control for both $\mathbf{X}$ and $\varepsilon_1$, player 1's choice $y_1$ is independent of $y_2$. The test is now more elaborate since we need to control not only for the observable covariates but also for the player's own unobservable (to the econometrician) shock. Following Aradillas-Lopez (2010), we add the following assumption:

***A-4*** *(Correlation structure) The joint distribution $G_{\varepsilon_1,\varepsilon_2}$ is such that a single parameter $\rho$ summarizes the correlation between $\varepsilon_1$ and $\varepsilon_2$.*

Under CPS, (8) and (9) can be written as[14]

$$\begin{aligned} y_1 &= \mathbb{1}\{h_1(X_1) + \alpha_1 \pi_2^*(\mathbf{X}, \varepsilon_1) - \varepsilon_1 \geq 0\} \\ &= \mathbb{1}\{\psi_1(\mathbf{X}, \varepsilon_1) \geq 0\}, \end{aligned} \tag{21}$$

and

$$y_2 = \mathbb{1}\{\psi_2(\mathbf{X}, \varepsilon_2) \geq 0\}. \tag{22}$$

[12] If the game has more than 2 players, we can add $\delta_2 y_3$ etc for each player since, under the null, only the $X$'s determine the decision.

[13] Another explanation for the rejection the null hypotheses described above could be the presence of market-level payoff shocks unobserved to the econometrician. In the next section we show that the test can be generalized to account for correlation across players unobservables.

[14] As before, the exact model is not important in terms of testing. The test works for any model that assumes the same information structure (i.e. CPS) and hence generates the same decision rule as in the second line below.

Thus, for an arbitrary value of $\rho$, the probability that both players enter is

$$\begin{aligned} & \Pr\left(y_1 = 1, y_2 = 1 | \mathbf{X}, \rho\right) \\ = & \int \mathbb{1}\left\{\psi_1\left(\mathbf{X}, \varepsilon_1\right) \geq 0\right\} \mathbb{1}\left\{\psi_2\left(\mathbf{X}, \varepsilon_2\right) \geq 0\right\} g_{\varepsilon_1, \varepsilon_2}\left(\varepsilon_1, \varepsilon_2; \rho\right) d\varepsilon_1 d\varepsilon_2, \end{aligned} \tag{23}$$

and the remaining probabilities can be defined accordingly.

Now consider testing the following null hypothesis:

$H_0$ : shocks are correlated but realizations are private information

$H_1$ : part of shocks are observed.

To make the test operational, we replace $\mathbb{1}\left\{\psi_1\left(\mathbf{X}, \varepsilon_1\right) \geq 0\right\}$ and $\mathbb{1}\left\{\psi_2\left(\mathbf{X}, \varepsilon_2\right) \geq 0\right\}$ in the objective function (e.g. likelihood) of the problem defined by the above equations with

$$\mathbb{1}\{\psi_1\left(\mathbf{X}, \varepsilon_1\right) + \delta_1 \varepsilon_2 \geq 0\}, \tag{24}$$

$$\mathbb{1}\left\{\psi_2\left(\mathbf{X}, \varepsilon_2\right) + \delta_2 \varepsilon_1 \geq 0\right\} \tag{25}$$

respectively. By doing this, we define a new hypothesis for player 1:

$$H_0' : \delta_1 = 0 \tag{26}$$

$$H_1' : \delta_1 \neq 0. \tag{27}$$

We can define a similar hypothesis for player 2 or even test for the joint event that both $\delta_1$ and $\delta_2$ are zero. The key point to notice is that rejection of $H_0'$ implies the rejection of $H_0$. That is, according to the correlated private shocks model, once we control for $\mathbf{X}$ and $\varepsilon_1$ in player 1's choice probability the remaining information contained on player 2's choice ($\varepsilon_2$) should not help predict player 1's choice. If it does, it means the information structure of the game is misspecified. Specifically, a player unobservables (from the econometrician's perspective) are at least partially observable by the other player.

## 3.3 Multiple Equilibria

Because recovering the structural form (i.e. the parameters) of the model is not our goal, but rather to test the different information structures, our test is "robust" to the problem of multiple equilibria. However, one important assumption we make is that the equilibrium selection rule is deterministic conditional on $\mathbf{X}$. To see why, consider an example of IPS. If there is only one equilibrium conditional on $\mathbf{X}$, we have

$$E(y_1 y_2|\mathbf{X}) = E(y_1|\mathbf{X})E(y_2|\mathbf{X}). \tag{28}$$

Now suppose that there are $J$ equilibria conditional on $\mathbf{X}$. Let $p_j(\mathbf{X})$ be the probability that the $j$-th equilibrium is played under a certain equilibrium selection rule. That is, $\sum_{j=1}^{J} p_j(\mathbf{X}) = 1$. Let $E_j$ be expectation operator when the $j$-th equilibrium is played. Then, we have

$$E(y_1 y_2|\mathbf{X}) = \sum_{j=1}^{J} p_j(\mathbf{X}) E_j(y_1 y_2|\mathbf{X}) \tag{29}$$

$$E(y_1|\mathbf{X})E(y_2|\mathbf{X}) = \left\{ \sum_{j=1}^{J} p_j(\mathbf{X}) E_j(y_1|\mathbf{X}) \right\} \left\{ \sum_{j=1}^{J} p_j(\mathbf{X}) E_j(y_2|\mathbf{X}) \right\}, \tag{30}$$

and clearly $E(y_1 y_2|\mathbf{X}) \neq E(y_1|\mathbf{X})E(y_2|\mathbf{X})$. Thus, a non-deterministic equilibrium selection rule breaks the conditional independence even if payoff shocks are entirely private information. This is the key intuition that de Paula and Tang (2011) use to test for the existence of multiple equilibria when they impose the independent private shocks assumption. Aradillas-Lopez and Gandhi (2011) characterize the conditions under which $E(y_1 y_2|\mathbf{X}) \geq E(y_1|\mathbf{X})E(y_2|\mathbf{X})$ holds, and use this moment inequality for inference of parameters of a certain class of models.[15]

Thus, one can understand our $\mu_i(\mathbf{X})$ and $\psi_i(\mathbf{X}, \varepsilon_i)$ functions as the reduced forms of the corresponding models provided the information structure is the same for the (unspecified) equilibrium selection rule and equilibrium assignments are deterministic conditional on common (public) information. Note that we do not assume that a single equilibrium is played in the data. We assume the existence of an equilibrium selection rule that depends on $\mathbf{X}$ and parameters, but not on any further randomness. That is, provided the equilibrium selection does not use more information, our semiparametric tests work for any model with the information structures we describe.

Alternatively, we could impose the assumption that the equilibrium selection rule is such that each player uses a different signal (independent of each other) to select an equilibrium. In this way, we could let the equilibrium selection depend on signals that the econometrician does not observe, and our testing procedure would be valid even in the presence of multiple equilibria.

[15] Specifically, Aradillas-Lopez and Gandhi (2011) consider ordered response games with incomplete information, which nest the entry game we consider in this paper. They derive a more general set of moment inequalities associated with the ordered response games.

### 3.4 Dependence between Observable Covariates and Payoff Shocks

Our test doesn't critically depend on the exogeneity assumption that the observable covariates and payoff shocks to players are independent. That is, we can allow $X_i$ and $\varepsilon_i$ to be correlated. For example, for the IPS information structure, a Bayesian-Nash equilibrium of this game is given by

$$y_1 = \mathbb{1}\{h_1(X_1) + \alpha_1 \pi_2^* - \varepsilon_1 \geq 0\}, \tag{31}$$

$$y_2 = \mathbb{1}\{h_2(X_2) + \alpha_2 \pi_1^* - \varepsilon_2 \geq 0\}, \tag{32}$$

where $(\pi_1^*, \pi_2^*)$ is a fixed point of $\varphi = (\varphi_1, \varphi_2) = \mathbf{0}$ with

$$\varphi_1(\pi_1, \pi_2) = \pi_1 - G_{\varepsilon_1|X_1}(h_1(X_1) + \alpha_1 \pi_2), \tag{33}$$

$$\varphi_2(\pi_1, \pi_2) = \pi_2 - G_{\varepsilon_2|X_2}(h_2(X_2) + \alpha_2 \pi_1). \tag{34}$$

Thus, the key result that the equilibrium probabilities are a function only of the observables $\mathbf{X}$ is still valid. In what follows, however, we keep the assumption that $X_i$ and $\varepsilon_i$ are independent for simplicity.

## 4 Properties of the Tests

While intuitive, it is not obvious that the tests we propose should have any power to discriminate alternative hypotheses. Since the tests we propose are standard t-tests, we expect them to behave well under the null. However, it is not clear whether the tests can reject the null when they should. In order to evaluate the power properties of our tests, in this section we perform a Monte Carlo study where we simulate the distribution of the test statistic under the relevant alternative hypotheses for different sample sizes and different values of the parameters controlling the departure from the null. As we show, the tests perform remarkably well for samples of even moderate sizes.

### 4.1 Simulation Design

For all the different models we present the basic parametrization we use is the following. We assume that $h(X_1) = \beta_1 X_1$ and $h(X_2) = \beta_2 X_2$. We set $\beta_1 = \beta_2 = 0.1$ and $\alpha_1 = \alpha_2 = -1.5$. The observable covariates $X_1$ and $X_2$ are randomly drawn from $U[2, 12]$. Each model is distinguished by the assumptions about the distribution of the unobservables $\varepsilon_1, \varepsilon_2$ as well as the specification of the information available to each player $\Omega$.

#### 4.1.1 Independent Private Shocks

We assume that the shocks $\varepsilon_1, \varepsilon_2$ are independent and that both follow standard normal distributions. For any draw $m$ of $(\mathbf{X}_m, \varepsilon_{1m}, \varepsilon_{2m})$ we form

$$y_{1m} = \mathbb{1}\{0.1X_{1m} - 1.5\pi^*_{2m}(X_1, X_2) - \varepsilon_{1m} \geq 0\} \quad (35)$$

$$y_{2m} = \mathbb{1}\{0.1X_{2m} - 1.5\pi^*_{1m}(X_1, X_2) - \varepsilon_{2m} \geq 0\}, \quad (36)$$

where $\pi^*_{1m}(X_1, X_2)$ and $\pi^*_{2m}(X_1, X_2)$ are the fixed point of

$$\pi_1 - \Phi(0.1X_{1m} - 1.5\pi_2) = 0 \quad (37)$$

$$\pi_2 - \Phi(0.1X_{2m} - 1.5\pi_1) = 0. \quad (38)$$

We calculate an equilibrium for each market as follows. Draw $X_{1m}, X_{2m}, \varepsilon_{1m}$ and $\varepsilon_{2m}$. We then find the equilibrium probabilities by finding the fixed point to (37) and (38).[16] To do so, we follow Aradillas-Lopez (2010) and start the fixed point search at $\pi_2 = 1$. Let $\pi_1^1$ be the solution to (37). Using $\pi_1^1$, let $\pi_2^1$ be the solution to (38). We iterate until we get $|\pi_1^k - \pi_1^{k+1}| < \epsilon$ and $|\pi_2^k - \pi_2^{k+1}| < \epsilon$ for sufficiently small $\epsilon$. Call the fixed point we obtain $\pi_1^*$ and $\pi_2^*$. Using these values, determine $(y_1, y_2)$ from the threshold crossing model given by (35) and (36). We calculate the equilibrium this way $M$ times.

#### 4.1.2 Correlated Private Shocks

In this case, we assume the shocks are distributed jointly normal:

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ & 1 \end{pmatrix}\right),$$

where, as a baseline, we set $\rho = 0.5$.

Calculating the fixed point for (10) and (11) is computationally demanding since, for given $\mathbf{X}$, we need to get a fixed point of functions $\pi_1^*(\mathbf{X}, \cdot)$ and $\pi_2^*(\mathbf{X}, \cdot)$.[17] To do so, we approximate (10) and (11) as follows. We first choose quadrature nodes $z_1, z_2, ..., z_{N_s}$ and quadrature weights $w_1, w_2, ..., w_{N_s}$ based on the Gauss-Chebyshev rule adapted to $(-\infty, \infty)$. For each $\mathbf{X}_m = \{X_{1m}, X_{2m}\}$, set $\pi_1^0(\mathbf{X}_m, \cdot) = 1$ and $\pi_2^0(\mathbf{X}_m, \cdot) = 0$.

[16] In general we do not have uniqueness of equilibrium in this setting (since we use normal distributions and both $\alpha_1$ and $\alpha_2$ are negative). Our choice is to simply use the first fixed point found. For the formal analysis of multiple equilibria in estimation of games of incomplete information, see Aradillas-Lopez (2010).

[17] As before, uniqueness of such a function is not guaranteed. In practice, we use the fixed point that is found first.

For all $\varepsilon_2 \in \{z_1, z_2, ..., z_{N_s}\}$, we update $\pi_1^1(\mathbf{X}, \varepsilon_2)$ using

$$\pi_1^{k+1}(\mathbf{X}_m, \varepsilon_2) \approx \sum_{s=1}^{N_s} \mathbb{1}\left\{0.1X_{1m} - 1.5\pi_2^k(\mathbf{X}_m, z_s) - z_s \geq 0\right\} \phi\left(z_s; \rho\varepsilon_2, 1-\rho^2\right) w_s, \tag{39}$$

where $\phi(\cdot; a, b)$ is the PDF of a normal distribution with mean $a$ and variance $b$. Likewise, for all $\varepsilon_1 \in \{z_1, z_2, ..., z_{N_s}\}$, we update $\pi_2^1(\mathbf{X}, \varepsilon_1)$ using

$$\pi_2^{k+1}(\mathbf{X}_m, \varepsilon_1) \approx \sum_{s=1}^{N_s} \mathbb{1}\left\{0.1X_{2m} - 1.5\pi_1^k(\mathbf{X}_m, z_s) - z_s \geq 0\right\} \phi\left(z_s, \rho\varepsilon_1, 1-\rho^2\right) w_s. \tag{40}$$

We then iterate the procedure until convergence.

Let $\pi_1^*(\mathbf{X}, \cdot) = \pi_1^{k+1}(\mathbf{X}, \cdot)$ and $\pi_2^*(\mathbf{X}, \cdot) = \pi_2^{k+1}(\mathbf{X}, \cdot)$ be the functions obtained from the fixed point algorithm described above. We then calculate $y_{1m}$ and $y_{2m}$ based on

$$y_{1m} = \mathbb{1}\{0.1X_{1m} - 1.5\pi_2^*(\mathbf{X}_m, \varepsilon_{1m}) - \varepsilon_{1m} \geq 0\} \tag{41}$$

$$y_{2m} = \mathbb{1}\{0.1X_{2m} - 1.5\pi_1^*(\mathbf{X}_m, \varepsilon_{2m}) - \varepsilon_{2m} \geq 0\} \tag{42}$$

for $m = 1, ..., M$.

#### 4.1.3 Partially Observable Shocks

In this case, we assume the shocks are distributed as

$$\begin{aligned} \varepsilon_1^o, \varepsilon_2^o &\sim N\left(0, \sigma_o^2\right) \\ \varepsilon_1^u, \varepsilon_2^u &\sim N\left(0, \sigma_u^2\right) \end{aligned}$$

and use the normalization $\sigma_o^2 + \sigma_u^2 = 1$. Notice that as $\sigma_o^2 \to 1$ all the random shocks become common knowledge, while as $\sigma_o^2 \to 0$ then the shocks become entirely private information. The data generating process is as follows: for market $m = 1, ..., M$ the equilibrium is given by

$$y_{1m} = \mathbb{1}\left\{0.1X_{1m} - 1.5\pi_2^*\left(X_{1m}, X_{2m}, \varepsilon_{1m}^o, \varepsilon_{2,m}^o\right) - \varepsilon_{1m}^o - \varepsilon_{1m}^u \geq 0\right\} \tag{43}$$

$$y_{2m} = \mathbb{1}\left\{0.1X_{2m} - 1.5\pi_1^*\left(X_{1m}, X_{2m}, \varepsilon_{1m}^o, \varepsilon_{2,m}^o\right) - \varepsilon_{2m}^o - \varepsilon_{2m}^u \geq 0\right\}, \tag{44}$$

where $\pi_1^*\left(X_1, X_2, \varepsilon_{1m}^o, \varepsilon_{2m}^o\right)$ and $\pi_2^*\left(X_1, X_2, \varepsilon_{1m}^o, \varepsilon_{2m}^o\right)$ are given by the solution to the following system of equations:

$$\begin{aligned} \pi_1 - \Phi_{\varepsilon_1^u}\left(0.1X_{1m} - 1.5\pi_2 - \varepsilon_{1m}^o\right) &= 0 \quad (45) \\ \pi_2 - \Phi_{\varepsilon_2^u}\left(0.1X_{2m} - 1.5\pi_1 - \varepsilon_{2m}^o\right) &= 0, \quad (46) \end{aligned}$$

where we obtain the equilibrium choice probabilities in a similar manner as the IPS case except that now we do it for a given $(\mathbf{X}, \varepsilon_1^o, \varepsilon_2^o)$.

## 4.2 Implementation

In order to implement estimation on our simulated samples we use series estimators for the payoff functions. We approximate $\mu_i\left(\mathbf{X}\right)$ $i = 1, 2$ with the polynomial:

$$\mu_i\left(\mathbf{X}\right) = \lambda_{0i} + \lambda_{1i}X_i + \lambda_{2i}X_i^2 + \lambda_{3i}X_j + \lambda_{4i}X_j^2 + \lambda_{5i}X_iX_j. \quad (47)$$

For $\psi_i\left(\mathbf{X}, \varepsilon_i\right)$ $i = 1, 2$ we use

$$\begin{aligned} \psi_i\left(\mathbf{X}, \varepsilon_i\right) &= \theta_{0i} + \theta_{1i}X_i + \theta_{2i}X_i^2 + \theta_{3i}X_j + \theta_{4i}X_j^2 + \theta_{5i}\varepsilon_i + \theta_{6i}\varepsilon_i^2 + \theta_{7i}X_iX_j \quad (48) \\ &\quad + \theta_{8i}X_i\varepsilon_i + \theta_{9i}X_j\varepsilon_i + \theta_{10i}X_iX_j\varepsilon_i + \theta_{11i}X_i^2\varepsilon_i + \theta_{12i}X_j^2\varepsilon_i. \end{aligned}$$

For any given test for a fixed number of markets $M$ and parameters of the model, we simulate 250 datasets. In our baseline simulation we set the number of markets at 250. As a check, when the data is generated under the null, we calculate the t-statistic for our auxiliary testing parameter in each of our 250 simulated datasets and confirm that it fails to reject the null around 95% of the time.

To evaluate the power of the tests, we need to know the distribution of the test statistics (or the 95% confidence interval) for $\hat{\delta}_i = 0$ under the alternative hypothesis. To do so, we use a nonparametric bootstrap procedure to obtain these distributions. That is, when the simulated datasets are generated under an alternative hypothesis (CPS, POS for the IPS null; POS for the CPS null) we bootstrap each simulated dataset 250 times in order to get the distribution of the test statistic. For each simulated dataset we then calculate the 95% confidence interval for the statistic and check whether it rejects the null. Finally we count the number of times this happens across our 250 simulated datasets. The percentage of the time the null is rejected under the alternative is the power of the test.

For each of the possible alternatives, we change $M$ and check how the power of the test changes with the number of observations. We also calculate the power under different values for $\rho$ when the alternative is CPS and different values for $\sigma_o^2$ under POS. We plot the power function against $M$ and $\rho$ (or $M$ and $\sigma_o^2$) while keeping everything else constant.

### 4.3 Monte Carlo Results

In this section we show the results of the Monte Carlo design we just described. As a first quick check, we first generate 250 datasets for each of our 3 baseline data generating processes. For each dataset, we then estimate the model under each of the 2 nulls we investigate including the auxiliary parameter ($\delta_i$) that our test is based on. In Table 1 we show the average estimate for $\delta_1$ as well as a 95% interval over the 250 simulations. Notice that these are not to be interpreted as confidence intervals and are just meant as a rough check for how well we expect our test to behave. As is clear from the table, when the data generating process and the null hypothesis coincide, the average estimate is very close to 0 with the interval centered around it. When the data generating process differs from the null (i.e. when the null is false) the average estimate is far from zero and the intervals barely contain zero (if at all).

To get a formal idea of how the tests perform, we then take each of the 250 simulated datasets and bootstrap them 250 times. Then, for each simulated dataset, we form the t-statistic by taking the estimated $\delta$ and dividing it over the standard error obtained from the bootstrapped distribution.[18] The last column of Table 1 counts the number of times that the null is rejected (i.e. the number of times the t-statistic is larger in absolute value than 1.96). The same pattern we see in our simple analysis without standard errors holds: the null is rejected (roughly) 5% of the time when the null is true and it is rejected between 54% and 96% of the time when it should be rejected. The power properties of the test are remarkably good even for datasets of the modest size (250 markets) we use in this baseline simulation. The fact that the test has a rejection rate of 54% when the data is generated from the POS model but the CPS is the null is surprising given the relatively small fraction of the variance of the shock we assume is partially observed by the agents for this particular simulation (25%).

Figures 1 through 3 give a better idea of the performance of the tests. In Figures 1 and 2 we show how the power of the test changes as we change the sample size when the model is estimated under the null of IPS and the data generating process is CPS with $\rho = 0.5$ (Figure 1) and when the data generating process is POS with $\sigma_o^2 = 0.25$ (Figure 2). The power calculation is done in the same way by generating 250 datasets and using 250 bootstrapped samples per dataset to calculate the rate of rejection. As we can see the simple t-test we propose has considerable power even for small samples of 50 observations. The test is able to reject

[18] Alternatively, we could form the 95% confidence interval for each dataset and check whether it contains zero. The results are essentially the same as when we form the t-statistic.

the null around 80% of the time under either alternative for sample sizes as small as 200 and it rejects almost 100% of the time for samples of 450 observations or more. Figure 3 performs the same calculation when we test whether the test rejects the null of CPS when the true data generating process is POS with $\sigma_o^2 = 0.25$ and $\sigma_o^2 = 0.45$. The power of the test is weakly increasing in the number of markets when $\sigma_o^2 = 0.25$. We speculate that this is due to simulation error. While the test is considerably less powerful in this case, the power is still good given the small sample sizes and small fraction of the opponent's shock that we assume is observed by the player. As expected, as we increase the proportion of the shock that is observable to the other player ($\sigma_o^2 = 0.45$), the test performs quite well.

In Figure 4 we show how the power function changes as we change not only the sample size but also $\rho$ for the case in which the data is generated from the CPS model and the null hypothesis is IPS. The power of the test is monotone on the sample size regardless of the degree of correlation between the shocks. Surprisingly the test looses power for high values of the correlation coefficient. Figure 5 repeats the exercise for the case in which the data comes from the POS model instead and we change both the sample size and $\sigma_o^2$. For this case, the test becomes monotonically more powerful for both increases in the sample size and/or increases in the fraction of the opponent's shock observed by the player. Finally, in Figure 6 we plot the power function for the case in which the data is generated from the POS model but the null is CPS. Although the power is not high when $\sigma_o^2$ is around 0.2 or 0.3, it increases quickly as $\sigma_o^2$ increases.

## 5 An Empirical Example

This section applies our simple test to data on entry in the US airline industry. We use this industry as our empirical example primarily because several influential papers have estimated the entry model using this data: e.g., Berry (1992) and Ciliberto and Tamer (2009). Both papers assume that payoff shocks are common knowledge. While our test cannot provide a direct support for the complete information assumption, we can test against another extreme of entirely private information. The rejection of the null hypothesis would be, at least, consistent with the assumption of complete information used in these papers. The second reason is that there is potentially a lot of firm-specific information that airline carriers observe about each other but that is not observed by the econometrician. Finally, the number of markets is large in this industry so that our unspecified reduced form function can be flexible when controlling for observable covariates.

Our data comes from the first quarter of 2006's Airline Origin and Destination Survey (DB1B). The market is defined as a route between the origin airport and the final destination airport, regardless of whether the passenger makes an intermediate stop or not. We assume that round trips are non-directional. That is, for example, a round trip ticket between ORD and JFK is the same no matter which airport is the origin or destination. We use the 50 largest airports in the U.S. and exclude several airport pairs.[19] The final dataset contains 1,212 markets. We focus on the 5 major US airlines (Delta, American, United, Southwest, and Northwest), which we simply call firm 1 through firm 5, respectively.

Each firm has two choices: enter or not enter. Let $y_i = 1$ if firm $i$ enters the market and 0 if it does not. The decision rule for firm $i$ in market $m$ is given by

$$y_{im} = \mathbb{1}\left( g_i \left(X_{im}, Z_m, D_m\right) + \alpha_i \sum_{j \neq i} \pi^i_{jm} - \varepsilon_{im} > 0 \right), \tag{49}$$

where $X_{im}$ is a firm specific measure of market potential, $Z_m$ is a measure for demand size of market $m$, and $D_m$ is a variable for cost of serving in market $m$. For $X_{im}$, we use the number of airports connected (by firm $i$) to either the origin or the final destination airport of market $m$. $Z_m$ and $D_m$ are defined as the product of city populations for two end point airports and the distance between the two end airports, respectively. $\pi^i_{jm}$ denotes firm $i$'s evaluation of the entry probability of firm $j$.

## 5.1 Testing Independent Private Shocks

Our first goal is to test the null hypothesis that shocks are independent private information. Under the null, the equilibrium beliefs are given by

$$\pi^{i*}_j = \pi^{i*}_j \left(X_1, ..., X_5, Z_m, D_m\right). \tag{50}$$

Following the analysis in the text, we estimate the following equation for firm 1:

$$y_{1m} = \mathbb{1}\left( \mu \left(X_{1m}, ..., X_{5m}, Z_m, D_m\right) + \sum_{j=2}^{5} \delta^1_j y_{jm} - \varepsilon_{im} > 0 \right). \tag{51}$$

We approximate the $\mu$ function as polynomial on the $X$'s, $Z_m$, $D_m$, and their interactions. First we assume $\varepsilon_{im}$ follows the standard normal distribution. The total number of parameters we estimate is 37. For simplicity, we test whether the $\delta_j$ are jointly zero:

$$\delta^1_2 = \delta^1_3 = \delta^1_4 = \delta^1_5 = 0 \tag{52}$$

[19] Several routes between several airports shouldn't be regarded as markets. For example, there is no flight between Chicago O'Hare and Chicago Midway, and also nobody recognizes it as a route for airplanes. Therefore, we exclude several pairs that have the same feature as this example.

The test statistic we use is the likelihood ratio test:

$$LR = 2(517.0 - 505.4) = 23.2, \tag{53}$$

which is larger than the critical value (13.3 at the 1% significance level).

If $\varepsilon$ does not follow the standard normal distribution, the model is misspecified and the auxiliary parameters may be biased. To alleviate this risk, we estimate the model under the same null hypothesis, assuming that $\varepsilon$ follows the mixture of two normal distributions. The total number of parameters is 39. The test statistic is

$$LR = 2(502.1 - 474.7) = 18.4, \tag{54}$$

which is larger than the critical value. To conclude, we reject the hypothesis that random shocks are entirely independent private information.

## 5.2 Testing Correlated Private Shocks

We next test the null hypothesis that shocks are correlated but private information. Under the null, the equilibrium beliefs are given by

$$\pi_j^{i*} = \pi_j^{i*}\left(X_1, ..., X_5, Z_m, D_m, \varepsilon_1\right). \tag{55}$$

We estimate the following equation:

$$\begin{aligned} & \Pr\left(y_1 = 1, ..., y_5 = 1 | X_1, ..., X_5, Z, D, \rho\right) \\ = & \int \prod_{i=1}^{5} \mathbb{1}\left\{\psi_i\left(X_1, ..., X_5, Z_m, D_m, \varepsilon_{im}\right) + \sum_{j \neq i} \delta_j^i \varepsilon_{jm} \geq 0\right\} g_\varepsilon\left(\varepsilon\right) d\varepsilon, \end{aligned} \tag{56}$$

where $g_\varepsilon$ denotes the density of the joint distribution of $(\varepsilon_1, ..., \varepsilon_5)$, which we assume is the multivariate normal distribution with a single parameter $\rho$.[20] The total number of parameters is 361 (340 in $\psi$, 20 $\delta$s, and $\rho$).

Again, we test whether all the $\delta_j^i$ are jointly zero. The test statistic of the likelihood ratio test is

$$LR = 2(2176.6 - 2086.1) = 181.0, \tag{57}$$

which is higher than the critical value of the chi-squared distribution with 20 degrees of freedom (37.6 at the 1% significance level). Therefore, we can conclude that even after controlling for exogenous correlation

[20] The diagonal elements of the variance-covariance matrix are normalized to one. The off-diagonal elements are all $\rho$.

between $\varepsilon_i$ and $\varepsilon_j$, the null hypothesis that payoff shocks are entirely private information is rejected. That is, airline companies partially (and potentially fully) observe competitors' payoff shocks not observable to the econometrician.

# 6 Conclusion

The literature on the estimation of games of incomplete information has paid close attention to the semiparametric and nonparametric identification and estimation of these games. However, in all cases, this is done under maintained assumptions about the information available to both players and the econometrician. As we show in this paper, a very simple specification test that allows one to check whether these assumptions are violated can be employed. Our test checks for violation of the conditional independence implied by an information structure. As we show, for the widely used examples of static entry games, the test can be implemented in a very simple and intuitive way. For the independent private shocks null hypothesis, the test consists of estimating a standard binary choice model which, under assumptions about the distribution of the shocks, is a standard problem. While simple, the test seems to have very good power properties even for samples of moderate size. The test of correlated private shocks, while not as powerful, still exhibits good power properties. Our simple empirical example on entry in the US airline industry shows that both the hypotheses of independent private shocks and of correlated private shocks are not supported by the data.

# References


[1] Aguirregabiria, V., Mira, P., 2007. Sequential Estimation of Dynamic Games. Econometrica 75(1), 1-54.

[2] Aradillas-Lopez, A., 2010. Semiparametric Estimation of a Simultaneous Game with Incomplete Information. Journal of Econometrics 157(2), 409-431.

[3] Aradillas-Lopez, A., Gandhi, A., 2011. Robust Inference in Ordered Response Games with Incomplete Information: Are Firms Strategic Substitutes? Working Paper, University of Wisconsin-Madison.

[4] Bajari, P., Hong, H., Krainer, J., Nekipelov, D., 2010. Estimating Static Models of Strategic Interactions. Journal of Business and Economic Statistics 28(4), 469-482.

[5] Berry, S., 1992. Estimation of a model of entry in the airline industry. Econometrica 60(4), 889-917.

[6] Bresnahan, T., Reiss, P., 1991. Entry and Competition in Concentrated Markets. Journal of Political Economy 99(5), 977-1009.

[7] Brock W., Durlauf, S., 2001. Discrete Choice with Social Interactions. Review of Economic Studies 68(2), 235-260.

[8] Ciliberto, F., Tamer, E., 2009. Market Structure and Multiple Equilibria in Airline Markets. Econometrica 77(6), 1791-1828.

[9] Collard-Wexler, A., 2010. Demand Fluctuations in the Ready-Mix Concrete Industry. Working Paper, New York University.

[10] Cunha, F., Heckman, J., Navarro, S., 2005. Separating Uncertainty from Heterogeneity in Life Cycle Earnings. Oxford Economic Papers 57(2), 191-261.

[11] de Paula, A., Tang, X., 2011. Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information. forthcoming in Econometrica.

[12] Grieco, P., 2010. Discrete Games with Flexible Information Structures: An Application to Local Grocery Markets. Working Paper, Northwestern University.

[13] Heckman, J., Navarro, S., 2004. Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models. Review of Economics and Statistics 86(1), 30-57.

[14] Mazzeo, M., 2002. Product choice and oligopoly market structure. RAND Journal of Economics 33(2), 1-22.

[15] Pesendorfer, M., Schmidt-Dengler, P., 2008. Asymptotic Least Squares Estimators for Dynamic Games. Review of Economic Studies 75(3), 901-928.

[16] Ryan, S., 2011. The Costs of Environmental Regulation in a Concentrated Industry. forthcoming in Econometrica.

[17] Seim, K., 2006. An Empirical Model of Firm Entry with Endogenous Product-Type Choices. RAND Journal of Economics 37(3), 619-640.

[18] Sweeting, A., 2009. The Strategic Timing of Radio Commercials: An Empirical Analysis Using Multiple Equilibria. RAND Journal of Economics 40(4), 710-742.

[19] Sweeting, A., 2011. Dynamic Product Positioning in Differentiated Product Industries: The Effect of Fees for Musical Performance Rights on the Commercial Radio Industry. Working Paper, Duke University.

Table 1: Average Auxiliary Parameter and Intervals

| Data Generated from: | Model Estimated Under the Null of: | Average Auxiliary Parameter | 95% Interval | Average t-statistic | Proportion of times the Null is Rejected |
|---|---|---|---|---|---|
| Independent Private Shocks | Independent Private Shocks | -0.0097 | [ -0.42040, 0.31408 ] | -0.06154 | 6.0% |
| Correlated Private Shocks $\rho$=0.5 | Independent Private Shocks | 0.6209 | [ 0.12528, 1.07074 ] | 2.63855 | 81.2% |
| | Correlated Private Shocks | 0.0069 | [ -0.41047, 0.46700 ] | 0.03217 | 5.6% |
| Partially Observable Shocks $\sigma^2_o$=0.25 | Independent Private Shocks | -0.7024 | [ -1.08509, -0.30726 ] | -3.59575 | 96.4% |
| | Correlated Private Shocks | -0.3484 | [ -0.69819, 0.10561 ] | -2.00783 | 54.4% |

Note: We generage 250 simulated datasets. For each dataset, we estimate the auxiliary parameter and then take an average across those 250 datasets as well as forming the 95% interval over the 250 simulations. We also bootstrap each simulated dataset 250 times to get the distribution. With this we form the standard error and the t-statistic. The last column contains the number of times the null hypothesis is rejected across datasets using this t-statistic.

Figure1: Power function
Null Hypothesis of Independent Private Shocks
Data Generated from Correlated Private Shocks with $\rho$=0.5

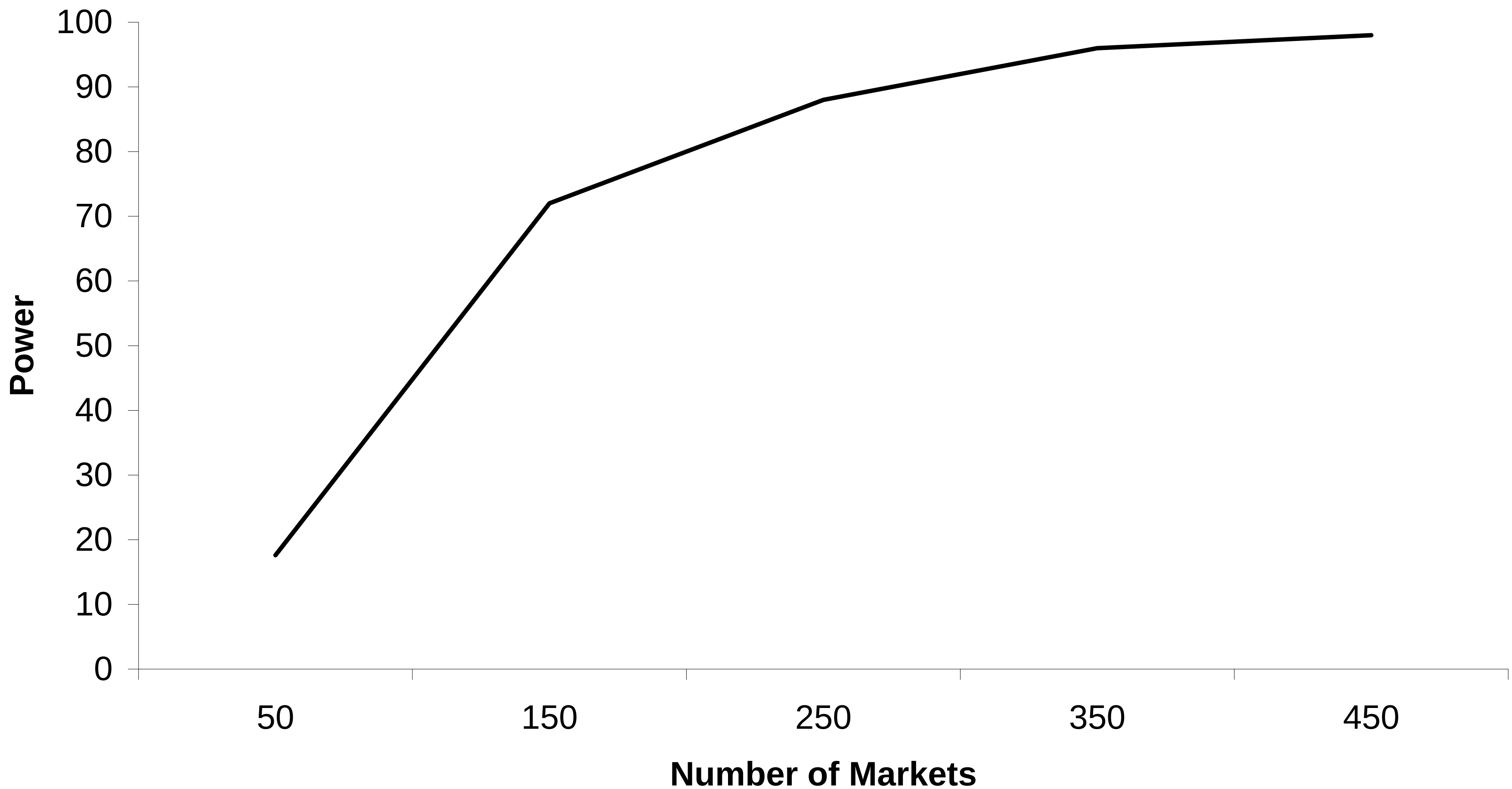


Figure 2: Power function
Null Hypothesis of Independent Private Shocks
Data Generated from Partially Observable Shocks with $\sigma^2_o$=0.25

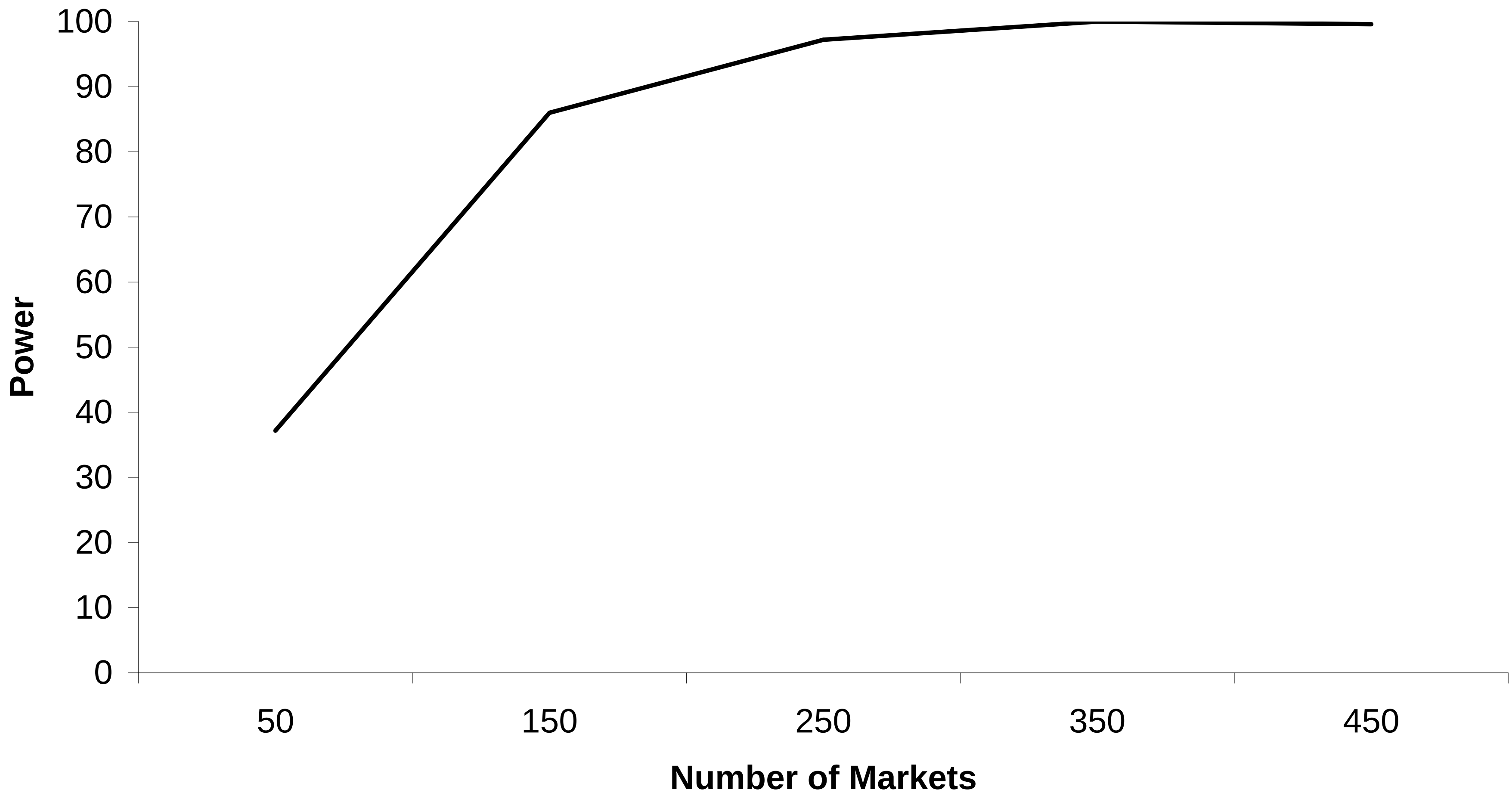


Figure 3: Power Function
Null Hypothesis of Correlated Private Shocks
Data Generated from Partially Observable Shocks

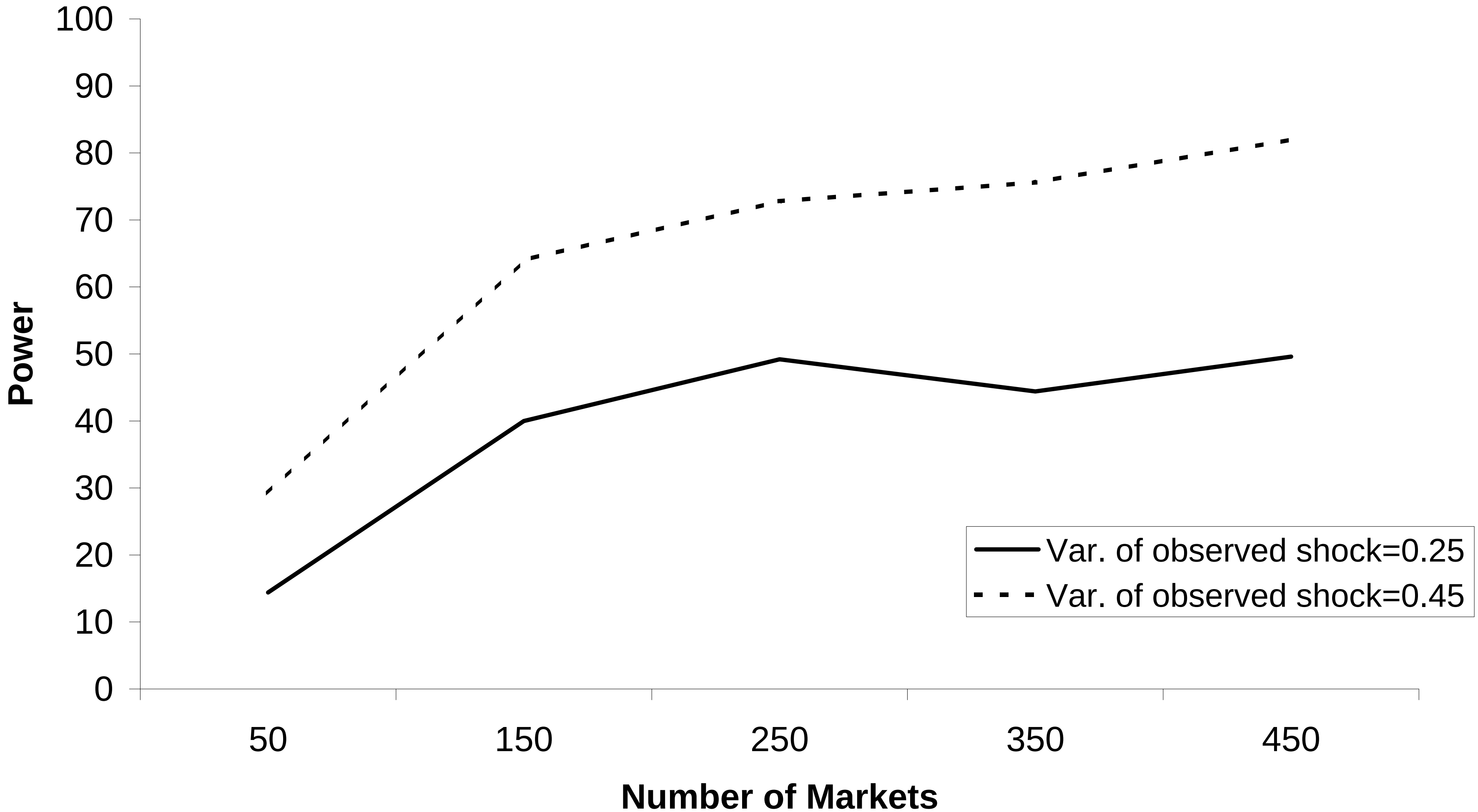
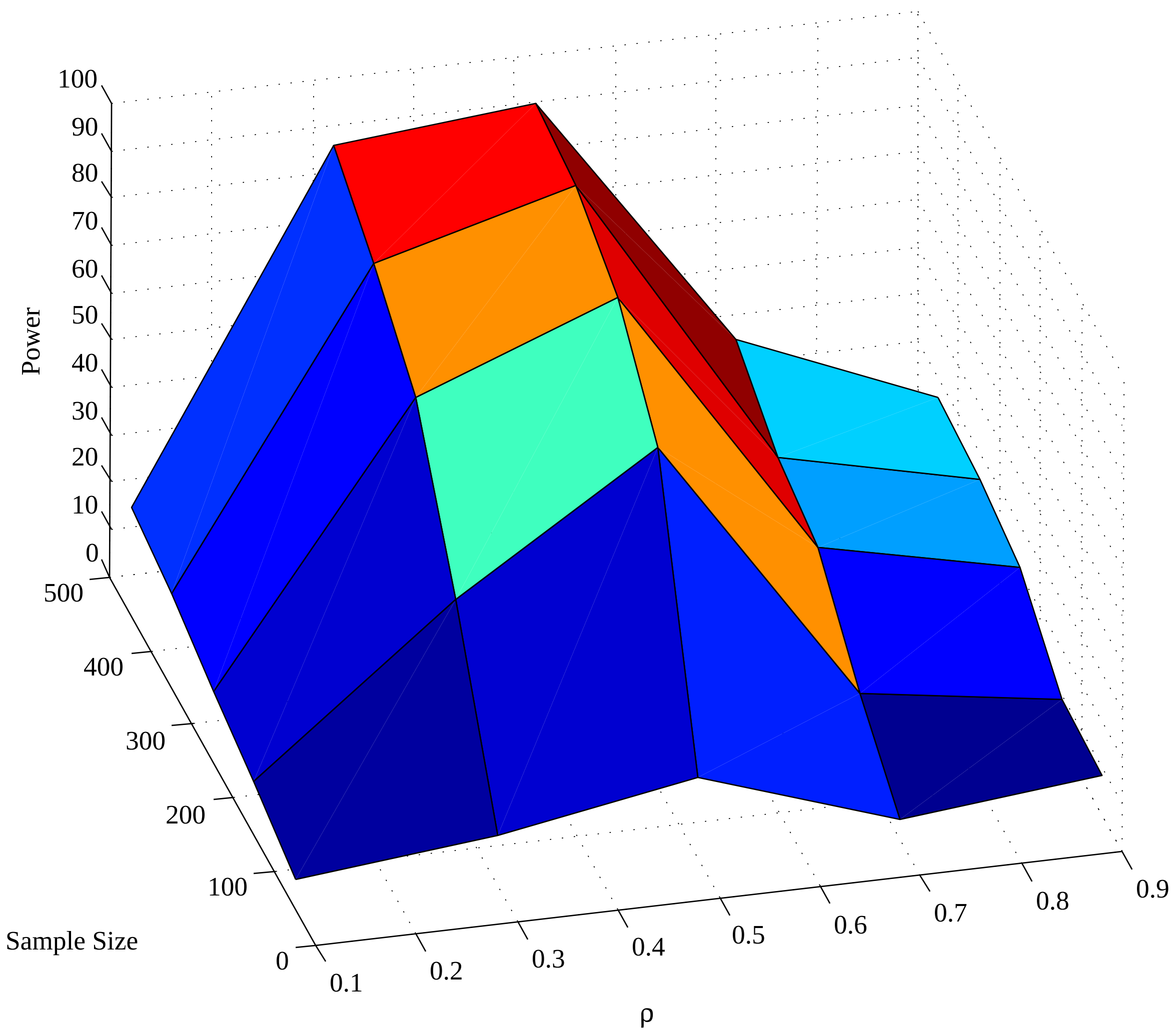
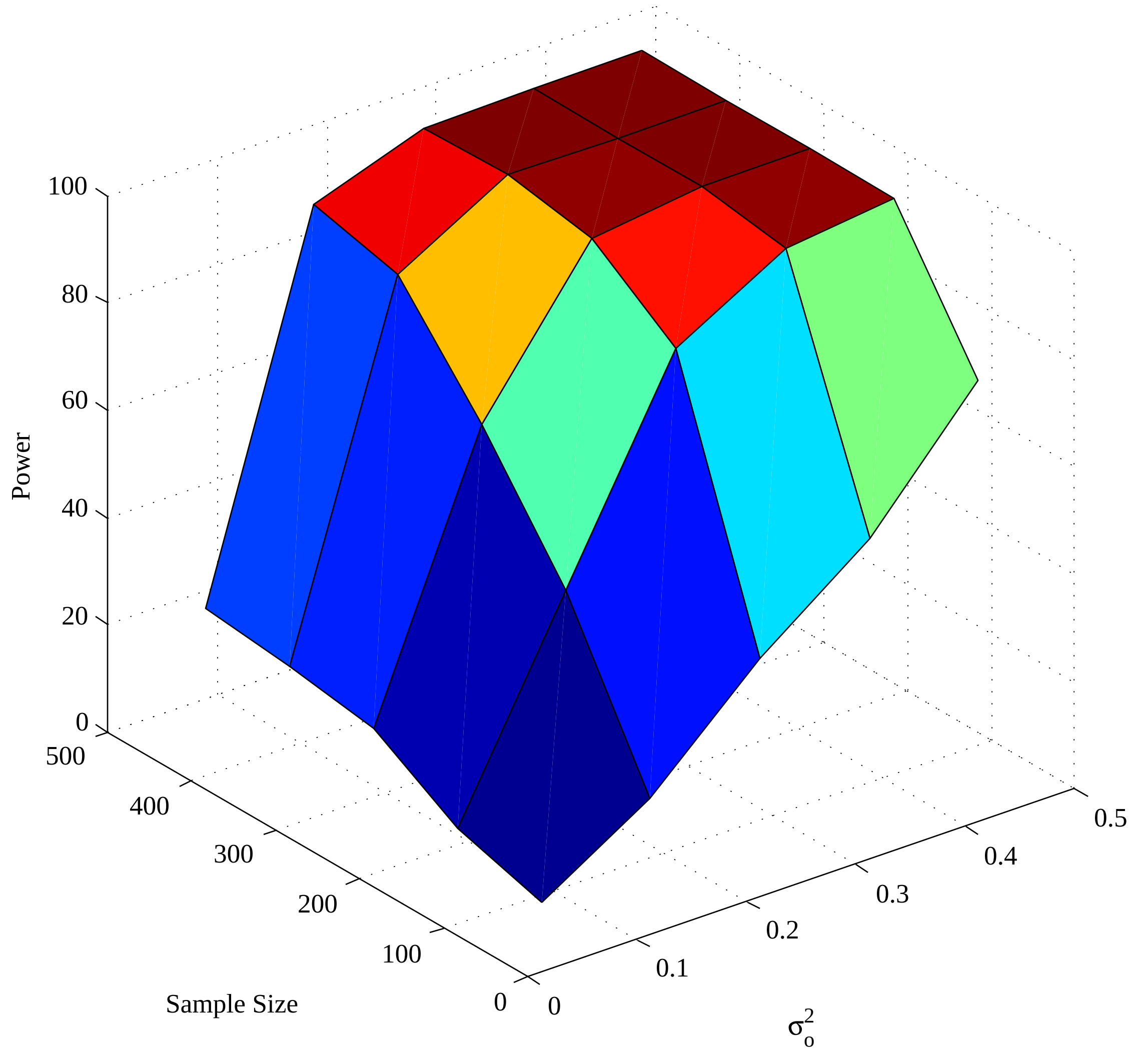


Figure 4: Power Function
Null Hypothesis: Independent Private Shocks
DGP: Correlated Private Shocks



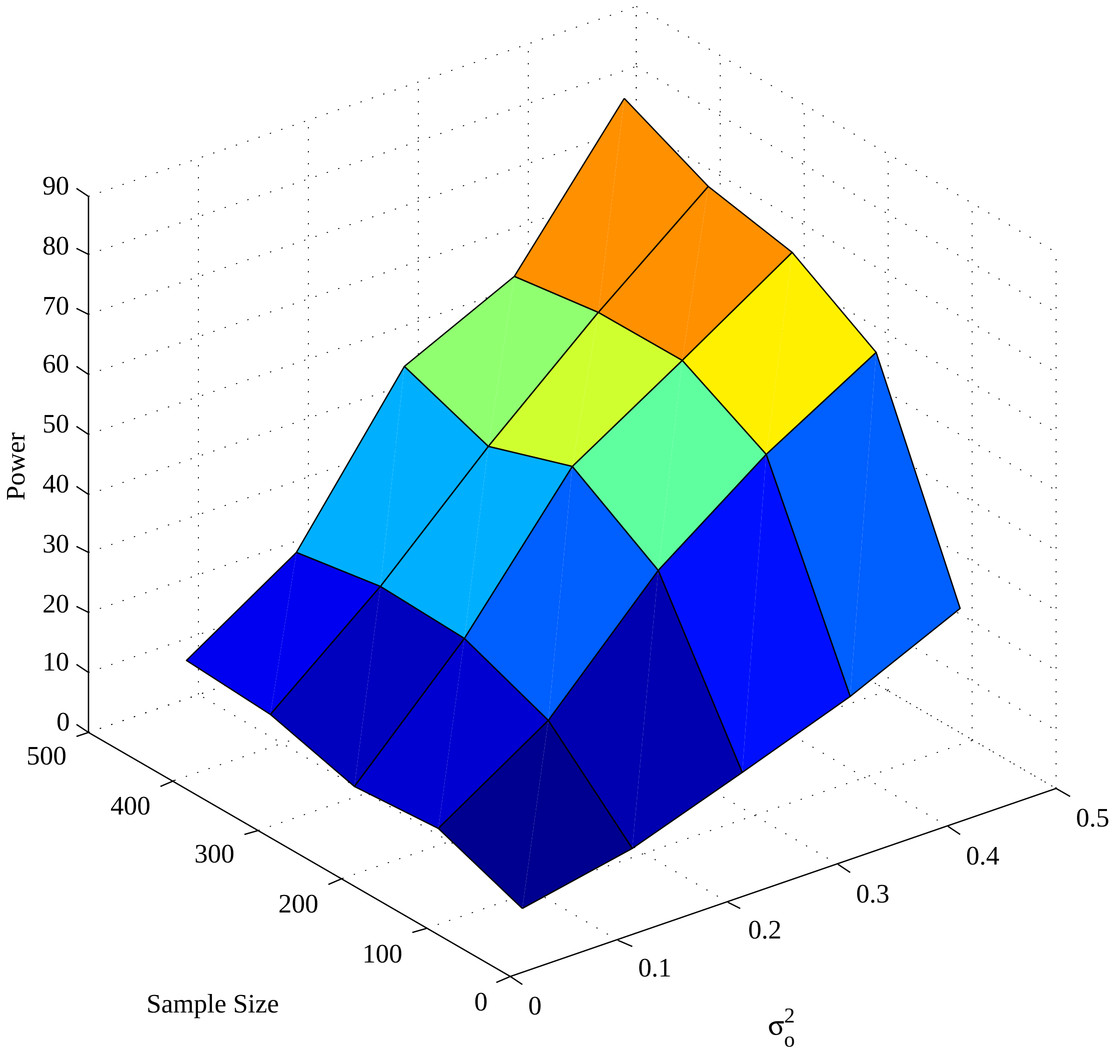

Note: We calculate the power for each pair of the number of markets and the correlation coefficient.

Figure 5: Power Function
Null Hypothesis: Independent Private Shocks
DGP: Partially Observable Shocks




Note: We calculate the power for each pair of the number of markets and the variance of observable shocks.

Figure 6: Power Function
Null Hypothesis: Correlated Private Shocks
DGP: Partially Observable Shocks




Note: We calculate the power for each pair of the number of markets and the variance of observable shocks.