ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

https://www.econometricsociety.org/

Econometrica, Vol. 88, No. 4 (July, 2020), 1725–1752

STATISTICAL INFERENCE IN GAMES

YUVAL SALANT
Kellogg School of Management, Northwestern University

JOSH CHERRY
Pinterest

Econometrica, Vol. 88, No. 4 (July, 2020), 1725–1752

STATISTICAL INFERENCE IN GAMES

YUVAL SALANT
Kellogg School of Management, Northwestern University

JOSH CHERRY
Pinterest

We consider statistical inference in games. Each player obtains a small random sample of other players' actions, uses statistical inference to estimate their actions, and chooses an optimal action based on the estimate. In a sampling equilibrium with statistical inference (SESI), the sample is drawn from the distribution of players' actions based on this process. We characterize the set of SESIs in large two-action games, and compare their predictions to those of Nash equilibrium, and for different sample sizes and statistical inference procedures. We then study applications to competitive markets, markets with network effects, monopoly pricing, and search and matching markets.

KEYWORDS: Statistical inference, bounded rationality, games, search and matching markets.

1. INTRODUCTION

THERE ARE MANY SITUATIONS in which an individual needs to decide whether to take an action whose value depends on the number of people taking it. For example, deciding whether to drive to work depends on road congestion; market thickness affects the likelihood that a job search is successful; and the value of adopting a new network product is influenced by other people's adoption decisions. In such settings, the individual has to estimate the number of people taking the action in order to determine its value.

The individual may act as a statistician when estimating this number. He may first obtain some data on other people's actions. For example, he may ask a few co-workers whether they drive to work. The individual may then use some form of statistical inference to estimate how many people take the action. For example, he may combine the data collected with some prior perception he has about road congestion, or perhaps estimate the most likely road congestion parameter to have generated the data. The individual can then use the estimate to determine the action's value and decide whether to take it.

This paper considers such statistical decision making in games. Each player obtains a small sample of other players' actions. To describe how players make inferences from the sample, we introduce the notion of an *inference procedure*, which is the analogue of an estimator in the statistics literature. An inference procedure assigns to every possible

Yuval Salant: y-salant@kellogg.northwestern.edu
Josh Cherry: josh.cherry@gmail.com
We thank Eduardo Azevedo, Joe Harrington, Emir Kamenica, Shengwu Li, Dmitry Lubensky, Sanket Patil, Neal Roese, Ariel Rubinstein, Ron Siegel, Ran Spiegler, and seminar participants in Bar-Ilan University, Ben-Gurion University, Brown University, Hebrew University, Northwestern University, Penn-State University, Tel-Aviv University, University of Pennsylvania, University of Toronto, University of Western Ontario, the Harvard-MIT joint seminar, the European Summer Symposium in Economic Theory on Advances in Behavioral Theory, the Advances in Behavioral Game Theory conference at University College London, the Applied Behavioral Theory conference at the University of Zurich, the BRIC conference, and the Recent Developments in Behavioral Economics and Mechanism Design conference in the University of Glasgow for helpful discussions and comments.

 https://doi.org/10.3982/ECTA17105

sample a player may obtain an estimate, which is a distribution over possible proportions of players taking the action. Estimates are assumed to relate to one another monotonically: fixing the sample size, a player puts a larger weight on more players taking the action as the number of observations in which people take the action increases. Examples of inference procedures include Bayesian updating from a non-degenerate prior, maximum likelihood estimation, and Beta estimation, among others. Inference procedures can also be used to describe various heuristics in information processing such as the tendency to underestimate sample variability (Tversky and Kahneman (1971)). Each player uses the estimate to choose an action that maximizes his payoff.

A sampling equilibrium with statistical inference (SESI) incorporates this statistical decision making procedure into Osborne and Rubinstein's (2003) sampling equilibrium. A SESI is a distribution of actions with the property that sampling from this distribution and using statistical inference to arrive at an optimal action results in the same distribution of actions. A SESI depends on players' sample sizes and inference procedures.

In a SESI, players have access to the equilibrium distribution of actions. Osborne and Rubinstein (1998, 2003) interpreted this assumption as reflecting a steady state of a dynamic process in which new players sample the actions of past players. We formalize this interpretation in Section 8.[1] Another interpretation is that of noisy retrieval of information from memory. A player may have been involved in the same or a similar interaction in the past and encoded in memory the distribution of actions. However, because of memory decay and forgetting, the player can only retrieve from memory a few imperfect signals about his past experience.[2] Put differently, instead of sampling other players' actions, a player retrieves information from his own memory. In this interpretation, Nash equilibrium (NE) reflects an ability to perfectly retrieve information from memory in contrast to the imperfect retrieval in a SESI.[3]

Section 3 studies SESIs in large games in which the action players consider taking has an idiosyncratic benefit and a cost that is increasing and convex in the number of players taking the action. We establish three main results, which serve as building blocks for studying applications in later sections. First, for any inference procedure and any sample size, there is a unique SESI. Second, the proportion of players who take the action in the SESI, which we call the SESI proportion, is smaller than the unique NE proportion for any sample size and any unbiased inference procedure with the property that the expected value of the estimate is the sample mean. The SESI proportion could be as small as half of the NE proportion for natural cost specifications. Third, the SESI proportion increases in the sample size under additional assumptions on the inference procedure. These three results extend, under certain assumptions, to cost specifications that are non-monotone in the number of players taking the action.

Section 4 studies two simple applications. The main one is to competitive markets. A unit mass of producers face a known demand function for their product. Each producer has to decide whether to produce a unit of the product at an idiosyncratic cost to be sold at the market price. The market price depends on the production decisions of all producers, and so each producer obtains data on the production decisions of other producers, and uses statistical inference to estimate the market supply.

[1]Sethi (2000) formalized this interpretation in a different context.

[2]A classic contribution to the study of memory decay and forgetting is Ebbinghaus (1885). (See Ebbinghaus (2003) for the English version.) A more recent contribution is Schacter (1999, 2002).

[3]A possible difficulty with the information retrieval interpretation is that sample observations are assumed to be independent. Vul and Pashler (2008) provided partial support for this independence assumption.

Because there is a unit mass of producers, their data—when aggregated over all producers—reflect the market supply accurately. But because each producer only obtains a small sample, some producers will underestimate and some will overestimate the market supply. This is in contrast to a standard competitive equilibrium in which each producer estimates the market supply accurately. We show that there is a unique SESI in this environment, and that the market supply in this SESI depends on the curvature of the inverse demand function. In particular, when the inverse demand is convex, there is overproduction and a lower market price in the unique SESI than in the standard competitive equilibrium.

Section 5 considers an application to markets with positive network effects. Each consumer needs to decide whether to adopt a new product at an idiosyncratic cost. Adoption benefits are positive and increasing in the number of adopters. Marginal adoption benefits initially increase and then decrease in the number of adopters resulting in an S-shaped benefit function. An S-shaped benefit function supports three NEs. We establish that there is a unique SESI in this environment for small sample sizes, and up to three SESIs for larger sample sizes. All the SESI proportions are between the smallest and largest NE proportions.

Section 6 considers an application to monopoly pricing when the value of a product to a consumer decreases in the number of other consumers who purchase the product. A leading example is status goods. We ask how the monopolist's optimal profit changes as consumers obtain more information on other consumers' purchase decisions. We establish that the monopolist's profit decreases in the amount of information consumers obtain, suggesting the monopolist wishes to reduce consumers' access to information about other consumers' demand.

Section 7 considers an application to search and matching markets. Workers in this market decide whether to engage in costly search for jobs, and firms decide whether to engage in costly search for workers. Following Petrongolo and Pissarides (2001), we postulate that a Cobb–Douglas matching function determines the likelihood of a worker-firm match as a function of the matching friction and market thickness on the workers side and the firms side.[4]

We depart from the rational expectations modeling assumption by assuming that workers (and similarly firms) do not know the thickness on the other side of the market and use statistical inference to estimate it. A SESI in this environment is a steady state in which workers obtain data from the distribution of firms' actions based on the firms' statistical inference, and firms obtain data from the distribution of workers' actions based on workers' statistical inference.

We establish that market thickness and employment are smaller in the unique SESI with positive employment than in the unique rational expectations equilibrium with positive employment. The gap in market thickness and employment can be proportional to the matching friction itself. This implies that statistical inference amplifies the effect of shifts in the matching friction on employment relative to the predictions of the standard rational expectations equilibrium.

Section 8 concludes with two comments. The first is about the steady-state interpretation of SESIs. We consider a dynamic process based on the primitives of Section 2, and establish that it converges to the unique SESI. The second comment is about heterogeneity in players' sample sizes and inference procedures. Such heterogeneity may arise due to

[4]Petrongolo and Pissarides (2001) wrote: "the stylized fact that emerges from the empirical literature is that there is a stable aggregate matching function of a few variables that satisfies the Cobb–Douglas restrictions with constant returns to scale in vacancies and unemployment."

differences in players' ability to obtain or process information, different time constraints, or different preferences over statistical inference methods. We establish that our uniqueness and comparison to NE results extend to environments with heterogeneous players. We also identify conditions on the profile of players' sample sizes and inference procedures under which our comparative statics results continue to hold.

This paper is related to the literature on belief formation in games that relaxes the NE assumption that players have correct beliefs about the equilibrium distribution of actions. Most relevant to the current paper is Osborne and Rubinstein's (2003) sampling equilibrium in which players obtain a sample of other players' actions and best-respond to the sample average. Osborne and Rubinstein (2003) analyzed the sampling equilibria of a voting model with two or three samples. Osborne and Rubinstein's (1998) $S(k)$-equilibrium is an earlier contribution in which players do not know the mapping from actions to payoffs, sample the payoff of each action k times, and choose the action with the highest payoff. Sethi (2000) studied the dynamic stability of $S(1)$-equilibria, and Mantilla, Sethi, and Cárdenas (2020) analyzed the efficiency and stability of $S(1)$-equilibria in public goods games. Spiegler (2006a,b) studied competition between firms that face consumers who sample their prices or another payoff-relevant parameter once.

There are other prominent approaches to belief formation in games. In a self-confirming equilibrium (Fudenberg and Levine (1993), Battigalli and Guaitoli (1997), Battigalli, Cerreia-Vioglio, Maccheroni, and Marinacci (2015)), players' beliefs cannot contradict the (possibly partial) feedback they obtain about payoffs and other players' behavior on the equilibrium path. In a cursed equilibrium (Eyster and Rabin (2005)), players may not account for the correlation between players' behavior and private information, and in a behavioral equilibrium (Esponda (2008)), players may also not account for the correlation between behavior and other payoff-relevant parameters. In an analogy-based equilibrium (Jehiel (2005)), players form coarse beliefs about the average behavior in analogy classes, which are a partition of other players' strategy sets. In a Berk–Nash equilibrium (Esponda and Pouzo (2016)), players entertain a collection of possible beliefs and use the one that fits best (according to a relative entropy criterion) the feedback they obtain. In Spiegler (2019), players use the maximum entropy criterion to form beliefs about payoff-relevant parameters after obtaining partial feedback on the correlations between them.

We make three contributions to this literature. First, we consider players who use statistical inference to form beliefs. Statistical inference is captured by the notion of an inference procedure, which is flexible enough to model a rich class of possible statistical inferences from data. A related contribution is Liang (2018) who studied whether players with different statistical learning rules converge to NE after observing the same long sequence of data on payoff-relevant parameters. In contrast, our emphasis is on statistical inference from small heterogeneous samples that leads to different predictions than NE. Another related contribution is Al-Najjar (2009) who studied statistical decision making in a single-person environment.

Our second contribution is to provide a comprehensive characterization of SESIs in large two-action games. Our ability to do so relies on the connection between SESIs and Bernstein polynomials that we discuss in detail in Section 3.1. Nöldeke and Peña (2016) and Peña, Lehmann, and Nöldeke (2014) used Bernstein polynomials in other game-theoretic settings to study NEs of voter participation games and evolutionary dynamics in two-action N-player games.

Our third contribution is to develop applications to competitive and network markets, monopoly pricing, and search and matching markets.

2. MODEL

This section presents the model. After describing the model primitives in Section 2.1, Section 2.2 discusses players' statistical decision making. Section 2.3 discusses the data generating process and the solution concept. Section 2.4 provides an example.

2.1. *Primitives*

There is a unit mass of players, and each of them has to decide whether to take action A or action B. The utility from action B is 0. The utility from action A is $u(\theta, \alpha) = \theta - f(\alpha)$, where θ is a player's idiosyncratic benefit from A, and $f(\alpha)$ is the cost incurred by a player taking action A if a proportion α of players take action A. The benefit θ is distributed uniformly on $[0, 1]$, and the function f is continuous, convex, and increasing with $f(0) \geq 0$ (i.e., f is a cost) and $f(1) \leq 1$ (i.e., the cost is weakly smaller than the maximal benefit of action A). Alternative cost specifications are considered in Section 3.6. A model with more than two actions is considered in the Supplemental Material (Salant and Cherry (2020)).

2.2. *Statistical Decision Making*

In order to decide which action to take, each player estimates the proportion α of players taking action A. The player obtains k independent observations from a Bernoulli distribution with probability of success α. Observing a success is interpreted as observing a player who takes action A, and a failure as observing a player who takes action B. The resulting sample is denoted by (k, z), where the integer $k \geq 1$ is the sample size and $z \in [0, 1]$ is the sample mean, that is, the proportion of successes in the sample.[5]

An inference procedure, which is the analogue of an estimator in the statistics literature, describes how a player makes inferences from the sample. It requires that the player puts more weight on more players taking action A as the sample mean increases.

DEFINITION: An *inference procedure* $G = \{G_{k,z}\}$ assigns a cumulative distribution function $G_{k,z}$, called an *estimate*, to every sample (k, z) such that the estimate $G_{k,\hat{z}}$ strictly first-order stochastically dominates the estimate $G_{k,z}$ when $\hat{z} > z$.

Following are a few examples of inference procedures.

EXAMPLE 1—Bayesian Updating: A player has a non-degenerate prior on α, and he uses Bayes' rule to update this prior based on the sample. By Proposition 1 in Milgrom (1981), the family of posteriors is an inference procedure.

EXAMPLE 2—Maximum Likelihood Estimation (MLE): A player uses the maximum likelihood method to estimate the most likely parameter to generate the sample, that is, the player solves for the α that maximizes $\alpha^{kz}(1-\alpha)^{k(1-z)}$. It is easy to verify that this α is z. The player treats this α as the proportion of players taking action A. Thus, the inference procedure is

$$G_{k,z}(\alpha) = 1_{\alpha \geq z} = \begin{cases} 0 & \alpha < z, \\ 1 & \alpha \geq z. \end{cases}$$

[5]The proportion z can take any value in $[0, 1]$ even though players only observe proportions of the form of j/k for $j = 0, 1, \ldots, k$. This is useful for comparative statics.

EXAMPLE 3—Beta Estimation: A player has "complete ignorance" about α. Such ignorance is often captured in the statistics literature by "Haldane's prior" (Haldane (1932), Zhu and Lu (2004)), which is the limit of the Beta(ϵ, ϵ) distribution as $\epsilon \to 0$.

When the player's sample includes only failures ($z = 0$) or successes ($z = 1$), the player concentrates his estimate on the sample mean similarly to MLE. When the sample includes both successes and failures, the player's estimate is the Beta$(zk, (1 - z)k)$ distribution. The mean of this estimate is z and its variance is $\frac{z(1-z)}{k+1}$.

To understand why a player with complete ignorance who observes the sample (k, z) may arrive at the estimate Beta$(zk, (1 - z)k)$, recall that the Beta distribution is a conjugate prior for Binomial distributions. Thus, a player with a Beta(a, b) prior on α who observes the sample (k, z) and uses Bayesian updating would have the posterior Beta$(a + zk, b + (1 - z)k)$. The sum $a + b + k$ of the two Beta parameters of the posterior can be divided into the weight $(a + b)$ placed on the prior and the weight k placed on the sample. For example, a uniform prior, which is a Beta$(1, 1)$ distribution, corresponds to putting a weight of 2 on the prior and the rest on the sample. In the current example, players are ignorant in the sense that they put no weight on a prior and base their estimate solely on the sample.[6]

EXAMPLE 4—Truncated Normal: A player believes that α is distributed according to a normal distribution truncated symmetrically around the mean. He estimates the mean to be z and the variance to be $\frac{z(1-z)}{k}$.

After obtaining the sample (k, z) and deriving the estimate $G_{k,z}$, a player best-responds to the estimate. That is, he takes the action A if and only if

$$\theta \geq F_{k,z},$$

where $F_{k,z}$ is the expected value of f under $G_{k,z}$:

$$F_{k,z} = \int_0^1 f(\alpha)\, dG_{k,z}(\alpha).$$

2.3. *Equilibrium*

Players obtain their sample from the distribution of actions based on players' statistical decision making. That is, players obtain data from a source that is representative of the parameter they wish to estimate. A sampling equilibrium with statistical inference describes the resulting solution concept.

DEFINITION: A *sampling equilibrium with statistical inference (SESI)* is a number $\alpha_{k,G} \in [0, 1]$ such that an $\alpha_{k,G}$ proportion of players take action A when each player obtains k independent observations from a Bernoulli distribution with probability of success $\alpha_{k,G}$, forms an estimate according to the inference procedure G, and best-responds to this estimate in choosing an action. We refer to $\alpha_{k,G}$ as the *SESI proportion of degree k with respect to the inference procedure G.*

[6]Beta estimation is not included in Example 1 because there does not exist a proper prior that, together with Bayesian updating based on the sample (k, z), generates a Beta$(zk, (1 - z)k)$ posterior.

A SESI embodies two procedural constraints on players' decision making. The first is informational: Players obtain data on the behavior of a small subset of players, for example, because of time or other constraints. If players had obtained data on the actions of all players, they would fully learn the equilibrium action profile, and the predictions of the model would coincide with those of NE in which players are assumed to know this profile. The second constraint is cognitive: Players use only their data and statistical inference to estimate the distribution of actions. If each player had investigated how other players make decisions, the player would arrive at the "correct" distribution of actions and best-respond to it, so the predictions of the model would again be identical to those of NE. The combination of the informational and cognitive constraints, however, implies different predictions than those of NE as we will see below.

2.4. *Example*

Assume that the utility of action A is $u(\theta, \alpha) = \theta - \alpha^4$, that is, $f(\alpha) = \alpha^4$, and that players use MLE. To solve for $\alpha_{1,\text{MLE}}$, the SESI proportion of degree 1 with respect to MLE, suppose that players obtain one observation from a Bernoulli distribution with success probability α. Then:

- With probability $1 - \alpha$, a player observes a failure. He estimates that no one takes action A, and hence takes this action if $\theta \geq f(0)$.
- With probability α, a player observes a success. He estimates that everybody takes action A, and hence takes this action if $\theta \geq f(1)$.

Thus, the proportion $\alpha_{1,\text{MLE}}$ has to satisfy $\alpha = (1 - \alpha) \cdot (1 - f(0)) + \alpha \cdot (1 - f(1))$. Rearranging, we obtain the equilibrium equation

$$1 - \alpha = (1 - \alpha) \cdot f(0) + \alpha \cdot f(1).$$

This linear equation has a unique solution at $\alpha_{1,\text{MLE}} = 1/2$.

To solve for $\alpha_{2,\text{MLE}}$, suppose that players obtain two observations. Then,

- With probability $(1 - \alpha)^2$, a player observes two failures. He estimates that no one takes action A, and hence takes this action if $\theta \geq f(0)$.
- With probability $2\alpha(1 - \alpha)$, a player observes one success and one failure. He believes that half of the population takes action A, and hence takes this action if $\theta \geq f(1/2)$.
- With probability α^2, a player observes two successes. He estimates that everybody takes action A, and hence takes this action if $\theta \geq f(1)$.

Thus, the proportion $\alpha_{2,\text{MLE}}$ has to solve the equilibrium equation

$$1 - \alpha = (1 - \alpha)^2 f(0) + 2\alpha(1 - \alpha) f(1/2) + \alpha^2 f(1),$$

which is quadratic in α and has a unique solution in [0, 1] at $\alpha_{2,\text{MLE}} \approx 0.60 > \alpha_{1,\text{MLE}}$.

Similarly, for sample size 3, the proportion $\alpha_{3,\text{MLE}}$ has to solve the equilibrium equation

$$1 - \alpha = (1 - \alpha)^3 f(0) + 3\alpha(1 - \alpha)^2 f(1/3) + 3\alpha^2(1 - \alpha) f(2/3) + \alpha^3 f(1),$$

which is cubic in α and has a unique solution in [0, 1] at $\alpha_{3,\text{MLE}} \approx 0.64 > \alpha_{2,\text{MLE}}$.

If players use Beta estimation as in Example 3 instead of MLE, then the SESI proportion changes for samples with two or more observations. For example, when the sample size is 2, a player who observes one success and one failure estimates that α is distributed according to the Beta(1, 1) distribution, and so the expected cost is $F_{2,1/2} = \int_0^1 f(z)\,dz$.

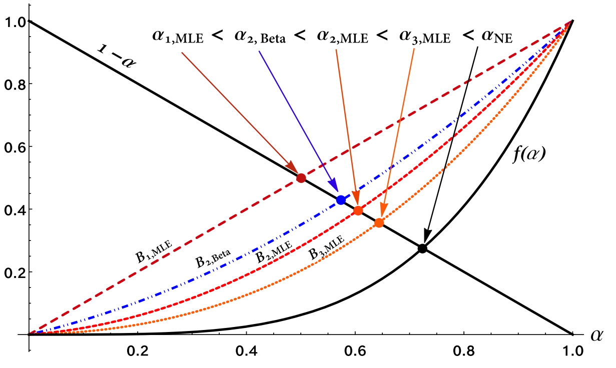

FIGURE 1.—Ranking of SESIs.

The equilibrium equation in this case is similar to the one that characterizes $\alpha_{2,\mathrm{MLE}}$ except for $F_{2,1/2}$ replacing $f(1/2)$:

$$1-\alpha=(1-\alpha)^2 f(0)+2\alpha(1-\alpha)F_{2,1/2}+\alpha^2 f(1).$$

This equation also has a unique solution in [0, 1] at $\alpha_{2,\mathrm{Beta}} \approx 0.57 < \alpha_{2,\mathrm{MLE}}$.

Figure 1 provides a graphical illustration. It plots in solid black the function $1-\alpha$ that appears on the left-hand side of all equilibrium equations. It plots in dashed colors the functions $B_{k,G}$, where $k \in \{1, 2, 3\}$ denotes sample size and $G \in \{\mathrm{MLE}, \mathrm{Beta}\}$ denotes the inference procedure. The functions $B_{k,G}$ correspond to the right-hand side of the various equilibrium equations. As can be seen from the figure, all the $B_{k,G}$ functions lie above f, which is plotted in solid black. The $B_{k,G}$ functions are ranked and their ranking implies the ranking of the corresponding SESIs.

3. EQUILIBRIUM CHARACTERIZATION

This section establishes several properties of SESIs, which will serve as building blocks for solving applications in later sections. These properties include existence and uniqueness (Section 3.2), how SESIs relate to NEs (Section 3.3), and how they change as the sample size or the inference procedure changes (Sections 3.4 and 3.5, respectively). Corresponding properties for alternative cost specifications are established in Section 3.6. The analysis relies on the theory of Bernstein polynomials presented in Section 3.1.

3.1. *SESIs and Bernstein Polynomials*

The main tool that we use in equilibrium analysis is the theory of Bernstein polynomials.

DEFINITION: For a function v defined on the closed interval [0, 1], the *kth-order Bernstein polynomial* of v is a function on [0, 1] defined by

$$\mathrm{Bern}_k(x; v) \equiv \sum_{j=0}^{k} \binom{k}{j} x^j (1-x)^{k-j} v(j/k).$$

That is, $\mathrm{Bern}_k(x; v)$ is a polynomial of degree k in x. It is the weighted average of the values of v at the $k+1$ points $\{0, \frac{1}{k}, \dots, \frac{k-1}{k}, 1\}$, where the weight assigned to

$v(j/k)$ is the probability of obtaining j successes in k independent observations from a Bernoulli distribution with success probability x. By definition, $\mathrm{Bern}_k(0; v) = v(0)$ and $\mathrm{Bern}_k(1; v) = v(1)$.

OBSERVATION 1: *A number $\alpha_{k,G}$ is a SESI proportion if and only if it solves the equation*

$$1 - \alpha = \mathrm{Bern}_k(\alpha; F_k), \tag{1}$$

where F_k is a function that assigns to every $z \in [0, 1]$ the value $F_{k,z}$, that is, $F_k(z) = F_{k,z}$.

PROOF: Suppose players obtain k independent observations from a Bernoulli distribution with success probability α and use the inference procedure G. The probability of observing j successes is $\binom{k}{j}\alpha^j(1-\alpha)^{k-j}$. Conditional on observing j successes, all players with $\theta \leq F_{k,j/k}$ take the action B, where $0 \leq F_{k,j/k} \leq 1$ because $0 \leq f(\alpha) \leq 1$. Thus, the fraction of players who observe j successes and take the action B is

$$\binom{k}{j}\alpha^j(1-\alpha)^{k-j}F_{k,j/k}.$$

Summing over j yields the total measure of players taking the action B, which is

$$\sum_{j=0}^{k}\binom{k}{j}\alpha^j(1-\alpha)^{k-j}F_{k,j/k} = \mathrm{Bern}_k(\alpha; F_k).$$

In equilibrium, this measure is equal to $1 - \alpha$. *Q.E.D.*

Observation 1 indicates that analyzing properties of SESIs is related to studying the behavior of Bernstein polynomials. The following properties of these polynomials, which they inherit from the function on which they operate, will be relevant in the analysis.

PROPERTY 1: If v increases in x, then $\mathrm{Bern}_k(x; v)$ increases in x.

PROPERTY 2: If v is convex, then $\mathrm{Bern}_k(x; v)$ is convex, and $\mathrm{Bern}_k(x; v) \geq \mathrm{Bern}_{k+1}(x; v) \geq v(x)$ for any x with strict inequality for $0 < x < 1$ if v is not linear.

Proofs of Properties 1 and 2 can be found in Phillips (2003). Another useful property that we prove directly is the following:

PROPERTY 3: Consider two inference procedures G and $\hat{G}$ such that $G_{k,z}$ is a mean-preserving spread of $\hat{G}_{k,z}$ for any sample (k, z). Then, $\mathrm{Bern}_k(\alpha; F_k) \geq \mathrm{Bern}_k(\alpha; \hat{F}_k)$ for any $\alpha \in [0, 1]$.

PROOF: Jensen's inequality and the convexity of f imply that $F_{k,j/k} \geq \hat{F}_{k,j/k}$ for any $0 \leq j \leq k$, which in turn implies the ranking of the corresponding Bernstein polynomials. *Q.E.D.*

3.2. *Equilibrium Existence and Uniqueness*

Our first result is about existence and uniqueness of SESIs.

THEOREM 1: *For any inference procedure and any sample size, there exists a unique SESI.*

PROOF: Fix an inference procedure G and sample size k, and consider Equation (1). The expression $1-\alpha$ on the left-hand side is strictly decreasing and continuous in α, it takes the value 1 at $\alpha=0$, and it takes the value 0 at $\alpha=1$.

The Bernstein polynomial $\mathrm{Bern}_k(\alpha; F_k)$ on the right-hand side of Equation (1) has the following properties. First, it is a continuous function in α on $[0,1]$. Second, it increases in α. This is because the first-order stochastic dominance of G implies that F_k is an increasing function, and so $\mathrm{Bern}_k(\alpha; F_k)$ is an increasing function (Property 1). Third, $0 \leq \mathrm{Bern}_k(0; F_k) < 1$ because $0 \leq F_k(0) < F_k(1) \leq 1$ and $F_k(0) = \mathrm{Bern}_k(0; F_k)$. Similarly, $0 < \mathrm{Bern}_k(1; F_k) \leq 1$.

Thus, $1-\alpha$ and $\mathrm{Bern}_k(\alpha; F_k)$ cross exactly once on $[0,1]$ implying existence and uniqueness. *Q.E.D.*

Figure 1 provides a graphical illustration. The Bernstein polynomials corresponding to the right-hand side of Equation (1) are plotted in dashed colors and are increasing in α. Each of them crosses the decreasing function $1-\alpha$ (plotted in solid black), which corresponds to the left-hand side of Equation (1), exactly once.

The proof of Theorem 1 does not rely on f being convex and thus extends to any continuous and increasing function. Section 3.6 identifies conditions on the inference procedure such that uniqueness also holds for any convex non-monotone f with $0 < f(0), f(1) < 1$.

3.3. *Relationship to Nash Equilibrium*

There is a unique NE proportion of players taking action A. To see this, consider a NE distribution of actions and let α_{NE} denote the proportion of players taking action A in this NE. All players with θ above (below) $f(\alpha_{\mathrm{NE}})$ strictly prefer to take action A (B) and thus the proportion α_{NE} has to satisfy $\alpha = 1 - f(\alpha)$ or

$$1-\alpha = f(\alpha). \tag{2}$$

Because f is increasing—and $1-\alpha$ is decreasing—in α, this equation has a unique solution on $[0,1]$.

To compare the SESI proportion, which depends on the inference procedure and the sample size, to the NE proportion, we restrict attention to unbiased inference procedures.

DEFINITION: An inference procedure G is *unbiased* if, for any sample (k, z), the expected value of the estimate $G_{k,z}$ is equal to the sample mean, that is,

$$\int_0^1 \alpha \, dG_{k,z}(\alpha) = z \quad \text{for any sample } (k, z).$$

In the statistics literature, estimates are often concentrated on a single point, and an estimator is unbiased if its expected value is equal to the underlying parameter. Unbiasedness in the current context implies a similar property. Fixing the sample size k and the underlying success probability α of the Bernoulli distribution, the expected value of

$G_{k,j/k}$ is j/k. This expected value is distributed according to a k-trial binomial distribution with success probability α, and so its expected value is α.

The inference procedures in Examples 2, 3, and 4 are unbiased. However, the Bayesian updating procedure in Example 1 is not. This is because whenever a player has a proper prior on α and he updates using Bayes' rule, the posterior mean of α depends on both the prior mean and the sample mean. In particular, the posterior mean cannot coincide with the sample mean for all samples.

THEOREM 2: *The SESI proportion is strictly smaller than the NE proportion for any unbiased inference procedure G, any sample size k, and any cost f that is not linear.*

PROOF: Fix a cost f, an unbiased inference procedure G, and a sample size k. The unique NE proportion solves Equation (2) and the unique SESI proportion solves Equation (1). Because the left-hand side of the equations is identical, it suffices to prove—in order to establish the result—that the continuous and increasing functions on the right-hand side of both equations are ranked such that $\mathrm{Bern}_k(\alpha; F_k) \geq f(\alpha)$ with strict inequality for $0 < \alpha < 1$ when f is not linear. This follows immediately from the equalities:

(i) $\mathrm{Bern}_k(\alpha; F_k) \geq \mathrm{Bern}_k(\alpha; f)$ for every $\alpha \in [0, 1]$, and

(ii) $\mathrm{Bern}_k(\alpha; f) \geq f(\alpha)$ for every $\alpha \in [0, 1]$ with strict inequality for $0 < \alpha < 1$ when f is not linear.

Inequality (ii) holds by Property 2. Inequality (i) holds by Property 3 because any estimate of an unbiased inference procedure is a mean-preserving spread of the corresponding MLE estimate, and the expected cost function for the MLE procedure is f itself. *Q.E.D.*

Figure 1 provides a graphical illustration. The Bernstein polynomials corresponding to the right-hand side of Equation (1) lie above the function f, which corresponds to the right-hand side of Equation (2). The ranking of the SESI and NE proportions follows because these proportions are the intersection points of the Bernstein polynomials and f, respectively, with $1 - \alpha$.

When the cost f is linear, the NE proportion and the SESI proportion coincide. This is because the linearity of f implies that $F_k = f$ and so $\mathrm{Bern}_k(\alpha; F_k) = \mathrm{Bern}_k(\alpha; f)$, and because the Bernstein polynomial of a linear function coincides with the function. This case is useful for highlighting another important difference between the two solution concepts, which relates to the selection of players who take each action.

In a NE, all players hold the same "correct" belief about the proportion of players taking each action. Therefore, there is a positive sorting of players to actions in the sense that if a player takes action A, then players with higher types also take action A. In a SESI, players' estimates differ from one another based on their sample. Players with a larger sample mean tend to take action A less than players with a smaller sample mean, leading to a weaker positive sorting than in a NE. It is even possible that there is no positive sorting of types to actions in a SESI. This happens, for example, in the SESI with sample size 1 in Section 2.4, in which players' actions depend only on their sample and not their type.

How different are the predictions of the unique SESI and the unique NE? We return to the example of Section 2.4 to demonstrate that the gap between the SESI and the NE proportions can be large. Recall that $\alpha_{1,\mathrm{MLE}}$ is 1/2 in this example. The NE proportion is the solution to $1 - \alpha = \alpha^4$, which is about 0.72. Consider now the function $f(\alpha) = \alpha^n$. As n increases, $\alpha_{1,\mathrm{MLE}}$ does not change. However, α_{NE} increases and converges to 1 as $n \to \infty$.

3.4. *Comparative Statics: Sample Size*

To evaluate how the SESI proportion changes as the sample size k increases, we make two additional assumptions on the inference procedure. The first is that players become more confident in their estimate as the sample size increases.

DEFINITION: An inference procedure G satisfies *noise reduction* if, for any two samples (k, z) and $(\hat{k}, z)$ such that $\hat{k} > k$, the estimate $G_{k,z}$ is a mean-preserving spread of $G_{\hat{k},z}$.

All the unbiased inference procedures in Section 2.2 satisfy noise reduction.

The second assumption is that fixing the sample size k, the inference procedure preserves the shape of the function f in the sense that the expected value of f as a function of the proportion of successes in the sample z is convex when f is convex.

DEFINITION: An inference procedure G *preserves shape* if the expected cost function F_k is convex when the cost function f is convex.

The maximum likelihood procedure trivially preserves shape because $F_k = f$. A sufficient condition for an unbiased inference procedure G to preserve shape is that, fixing k, the densities of the estimates $G_{k,z}(\alpha)$ are totally positive of degree 3 in (z, α) (cf. Jewitt (1988)).[7] The inference procedures in Examples 3 and 4 preserve shape because they belong to the exponential family, and exponential family densities are totally positive of every degree.

THEOREM 3: *If the inference procedure is unbiased, satisfies noise reduction, and preserves shape, then the SESI proportion of degree k is strictly smaller than the SESI proportion of degree $k+1$ when f is not linear.*

PROOF: Fix an inference procedure G. By Observation 1, it suffices to prove that for all $\alpha \in [0, 1]$,

$$\mathrm{Bern}_k(\alpha; F_k) \geq \mathrm{Bern}_{k+1}(\alpha; F_{k+1})$$

with strict inequality for $0 < \alpha < 1$. We establish this inequality in two steps.

STEP 1: $\mathrm{Bern}_k(\alpha; F_k) \geq \mathrm{Bern}_k(\alpha; F_{k+1})$. The convexity of f together with noise reduction imply by Jensen's inequality that $F_k(z) \geq F_{k+1}(z)$ for any $z \in [0, 1]$. The inequality follows.

STEP 2: $\mathrm{Bern}_k(\alpha; F_{k+1}) \geq \mathrm{Bern}_{k+1}(\alpha; F_{k+1})$ with strict inequality for $0 < \alpha < 1$. Because the inference procedure is shape preserving, F_{k+1} is convex. The inequality now follows from Property 2. *Q.E.D.*

Figure 1 provides a graphical illustration. Fixing the inference procedure, the Bernstein polynomials are ranked according to the sample size, and this ranking implies the ranking of the corresponding SESI proportions.

[7]A function $h(x, y)$ is totally positive of degree 3 if, for any $x_1 < x_2 < x_3$ and $y_1 < y_2 < y_3$, the matrix $(h(x_i, y_j))$ has a non-negative determinant for every minor of size ≤ 3. Total positivity of degree 2 is the monotone likelihood ratio property and is implied by total positivity of degree 3. Intuitively, total positivity of degree 3 ensures that the likelihood ratios increase sufficiently quickly to preserve convexity under integration.

Theorem 3 establishes that the SESI proportion gets closer to the NE proportion as the sample size increases. To obtain convergence to the NE proportion, we consider consistent inference procedures.

DEFINITION: An inference procedure G is *consistent* if, for any z, the sequence of estimates $\{G_{k,z}\}_{k=1}^{\infty}$ converges in probability to the distribution $1_{\alpha\geq z}$ that puts a unit mass on z.

All the inference procedures discussed in Section 2 are consistent.

OBSERVATION 2: *The SESI proportion converges to the NE proportion for any consistent inference procedure.*

PROOF: Fix a consistent inference procedure G. For any proportion of successes z, $F_{k,z} \xrightarrow[k\to\infty]{} f(z)$ because $G_{k,z} \xrightarrow[k\to\infty]{} 1_{\alpha\geq z}$. Moreover, the kth-order Bernstein polynomial of any continuous function converges to function as k tends to infinity (see, e.g., Phillips (2003) for a proof). Thus, the right-hand side of Equation (1) converges to the right-hand side of Equation (2), implying that the SESI proportion converges to the NE proportion. *Q.E.D.*

Intuitively, consistency implies that, fixing the sample mean z, players' estimates converge to z as the sample size k tends to infinity. Moreover, as k tends to infinity, the sample mean z, treated as a random variable, converges in probability to the underlying parameter α. Thus, players become fully informed about α as k tends to infinity similarly to NE.

3.5. *Comparative Statics: Inference Procedure*

The predictions of all unbiased inference procedures coincide when players obtain a single observation. This is because unbiasedness implies that a player who observes a success must concentrate his estimate on $\alpha = 1$, and a player who observes a failure must concentrate his estimate on $\alpha = 0$. However, two unbiased inference procedures may differ in how they estimate α when players obtain two or more observations.

DEFINITION: The inference procedure G is *more dispersed* than the inference procedure $\hat{G}$ if $G_{k,z}$ is a mean-preserving spread of $\hat{G}_{k,z}$ for any sample (k, z).

Fix a sample size k. If G is more dispersed than $\hat{G}$, then $\text{Bern}_k(\alpha; F_k) \geq \text{Bern}_k(\alpha; \hat{F}_k)$ by Property 3, and hence by Observation 1, the SESI proportion with respect to G is smaller than the SESI proportion with respect to $\hat{G}$. Thus, more dispersed inference implies a smaller proportion of players taking action A.

MLE is the least dispersed procedure among all unbiased procedures. Thus, for any given sample size, MLE provides the closest prediction to NE among all unbiased inference procedures. This is because, by Theorem 2, the SESI proportions of all unbiased inference procedures are smaller than the NE proportions.

3.6. *Alternative Cost Specifications*

The results of Sections 3.2–3.5 extend to additional cost specifications. We consider two extensions. The first extension is to cost functions that are increasing and concave. Theorem 1 continues to hold in this case as it relies only on the cost being increasing. The comparison between the SESI proportion and the NE proportion in Theorem 2 is reversed, that is, the SESI proportion is larger than the NE proportion, because the direction of the two relevant inequalities is reversed. As the sample size increases, the SESI proportion decreases because the relevant inequalities in Theorem 3 are also reversed. Finally, the SESI proportion is larger for a more dispersed inference procedure because the direction of the relevant Jensen inequality in Property 3 is reversed.

The second extension is to cost functions that are convex but not monotone. (Establishing analogous results for concave non-monotone functions follows the discussion in the previous paragraph.)

THEOREM 4: *Fix a strictly convex cost function f with $0 < f(0), f(1) < 1$ and an unbiased inference procedure G that preserves shape and satisfies noise reduction. Then, there exists a unique SESI for any sample size. The SESI proportion is smaller than the NE proportion and it increases in the sample size k.*

The proof of Theorem 4 is analogous to the proofs of Theorems 1–3. It appears in the Appendix, which contains all proofs that do not appear in the main text.

4. TWO EXAMPLES

This section incorporates statistical inference and the analysis of Section 3 into the theory of competitive markets (Section 4.1) and environments with a coordination dilemma (Section 4.2).

4.1. *Statistical Inference in a Competitive Market*

A unit mass of producers face a known inverse demand function $P(Q)$ for a good.[8] Each producer has to decide whether to produce a unit of the good at an idiosyncratic cost θ to be sold at the market price. The market price depends on the market supply Q, and so each producer needs to estimate Q in order to decide whether to produce.

In a NE, producers best-respond to the same correct belief about the market supply Q_{NE}. Producers may arrive at this belief using the following thought process: The market supply has to be equal to the proportion of producers who decide to produce given this market supply, that is, all producers with θ below the market price. Hence, Q_{NE} has to solve $Q = P(Q)$.

On the other hand, when producers act as statisticians, each of them obtains data on the production decisions of a few producers, uses statistical inference to estimate the market supply and the market price, and makes a decision based on the estimate. In a SESI, producers obtain their data from a Bernoulli distribution with success probability

[8]The function $P(Q)$ summarizes a demand model with a unit mass of consumers whose valuations are distributed on the interval [0, 1] with a positive density function. Because consumers' valuations are between 0 and 1, we can assume without loss of generality that $P(Q)$ is between 0 and 1.

governed by the market supply. Hence, the market supply Q_{SESI} solves

$$Q = \sum_{j=0}^{k} \binom{k}{j} Q^j (1-Q)^{k-j} P_{k,j/k},$$

where $P_{k,j/k}$ is the expected market price after obtaining the sample $(k, j/k)$ and forming the estimate $G_{k,j/k}$ of the market supply Q.

To apply the results of Section 3, we label the action "don't produce" as action A. Its benefit is θ and its cost is the foregone revenue from producing. Letting $N = 1 - Q$ denote the proportion of producers who do not produce, this cost is $f(N) = P(1 - N)$. Thus, the utility of not producing becomes $\theta - f(N)$ and the utility of producing becomes 0. Because P decreases in Q, f increases in N. And by definition, the functions f and P have the same convexity properties. The following result is a corollary of Theorem 1.

COROLLARY 1: *For any inference procedure and any sample size, there exists a unique SESI market supply* Q_{SESI}.

Because there is also a unique NE market supply, we can apply Theorem 2 and the market clearing condition to obtain the following result.

COROLLARY 2: *Fix an unbiased inference procedure. If the inverse demand function is convex in volume, then there is over-supply of the good and a lower market price in the SESI than in the NE. On the other hand, if the inverse demand function is concave in volume, then there is under-supply of the good and a higher market price in the SESI than in the NE.*

4.2. *Statistical Inference and the Dilemma of Limited Coordination*

This example is a variant of the El Farol Bar Problem (Arthur (1994)).

There is a unit mass of individuals and each of them needs to decide whether to go to a bar that offers entertainment on a Thursday night. The idiosyncratic travel cost to the bar is θ. The positive benefit $f(\alpha)$ of going to the bar depends on the proportion α of individuals going to the bar. As α increases, the marginal benefit f' decreases: it is positive up to $\bar{\alpha} < 1$, after which it is negative because of overcrowding. Thus, this is a coordination problem in which coordination benefits are maximized at $\bar{\alpha}$ and the benefit function f is positive, non-monotone, and concave.

A NE coordination level α_{NE} has the property that all individuals with $\theta \le f(\alpha_{\text{NE}})$ go to the bar, so α_{NE} has to solve $\alpha = f(\alpha)$. Despite the non-monotonicity of f, there a unique NE because the concave function f crosses the linear function α exactly once.

In a SESI, an individual who obtains the sample (k, z) goes to the bar if $\theta \le F_{k,z}$, where $F_{k,z}$ is the expected value of f according to the estimate $G_{k,z}$. Thus, the coordination level in a SESI solves $\alpha = \sum_{j=0}^{k} \binom{k}{j} \alpha^j (1-\alpha)^{k-j} F_{k,j/k}$.

Similarly to the analysis of Section 4.1, we need to label the action "don't go to the bar" as action A in order to apply Theorem 4 and obtain the following result.

COROLLARY 3: *Fix an unbiased inference procedure that preserves shape and satisfies noise reduction. Then, for any sample size, there exists a unique SESI. The coordination level in the SESI is smaller than in the unique NE, and it increases in the sample size.*

5. STATISTICAL INFERENCE IN NETWORK MARKETS

Comparing the SESI and NE proportions in Sections 3 and 4 relied on f being convex or concave. There are settings in which neither holds. For example, in the context of costly product adoption in a market with positive network effects, adoption benefits may be S-shaped: they may have increasing returns up to a certain level of users after which the returns may be decreasing. This section considers such a setting.

There is a unit mass of consumers and each of them needs to decide whether to adopt a new product at an idiosyncratic cost θ. The adoption benefit f is S-shaped: it is positive and increasing in the number of adopters, convex up to an inflection point $i \in (0, 1)$, and then concave. Suppose that f crosses the 45° line three times, $f(0) > 0$, and $f(1) < 1$.[9] And for tractability, suppose consumers use MLE to estimate the proportion of adopters.

Similarly to the analysis of Section 4, a SESI proportion has to solve $\alpha = \mathrm{Bern}_k(\alpha; f)$, where Bern_k operates on f because consumers use MLE. Because $\mathrm{Bern}_k(0; f) = f(0) > 0$ and $\mathrm{Bern}_k(1; f) = f(1) < 1$, $\mathrm{Bern}_k(\alpha; f)$ crosses the 45° line at least once and thus there is at least one SESI. There are three NEs because an NE proportion has to solve $\alpha = f(\alpha)$ and f crosses the 45° line three times.

The exact number of SESIs depends on the sample size and the variation in f close to 0 and 1. For sample size 1, there is a unique SESI because the SESI proportion is the intersection of the line $\alpha f(1) + (1 - \alpha)f(0)$ with the 45° line and $f(0) > 0$.

For larger sample sizes, the variation in f at the point j/k is $\delta_j^k = \Delta_{j+1}^k - \Delta_j^k$, where $\Delta_m^k = f(\frac{m}{k}) - f(\frac{m-1}{k})$. Thus, δ_j^k is positive (negative) if the line connecting $f(\frac{j-1}{k})$ and $f(\frac{j+1}{k})$ is above (below) $f(\frac{j}{k})$.

THEOREM 5: *There is a unique SESI of degree $k \geq 2$ if $\delta_1^k \leq 0$ or $\delta_{k-1}^k \geq 0$. Otherwise, there are at most three SESIs. Let $\bar{k}$ be the largest k such that either $\delta_1^k \leq 0$ or $\delta_{k-1}^k \geq 0$. Then, there is a unique SESI of degree $k \leq \bar{k}$.*

The intuition for the result is that $\delta_{k-1}^k \geq 0$ guarantees that the concave part of f does not affect the curvature of its Bernstein polynomial, which is convex and therefore intersects the 45° line once. Similarly, $\delta_1^k \leq 0$ guarantees that the convex part of f does not affect the curvature of its Bernstein polynomial, which is concave and therefore intersects the 45° line once. When $\delta_1^k > 0$ and $\delta_{k-1}^k < 0$, the Bernstein polynomial is S-shaped and may therefore cross the 45° line up to three times. Because $\delta_1^k \leq 0$ implies $\delta_1^{k-1} \leq 0$ and $\delta_{k-1}^k \geq 0$ implies $\delta_{k-2}^{k-1} \geq 0$, similar conclusions hold for smaller sample sizes.

The top panel of Figure 2 provides a graphical illustration. There are three NEs because f (in solid black) crosses the 45° line (also in solid black) three times. There is a unique SESI of degree 1 because the curve B_1 (in dashed blue), which denotes the first-order Bernstein polynomial of f, is linear. There is also a unique SESI of degree 2 because $\delta_1^2 > 0$ and hence B_2 (in dashed-dotted red) is convex. For sample sizes $k \geq 3$, $\delta_1^k > 0$ and $\delta_{k-1}^k < 0$ and hence the corresponding Bernstein polynomials are S-shaped. For $k = 3, 4$, the Bernstein polynomials (not depicted in the figure) cross the 45° line once despite being S-shaped and so there is a unique SESI. For larger sample sizes, the S-shaped Bernstein polynomial crosses the 45° three times, implying there are three SESIs. This is depicted in the figure for the curve B_5 (in dotted green). Note that the SESI proportions are between

[9]This assumption guarantees that there are three NEs and that the NE proportions are in the interior of $[0, 1]$.

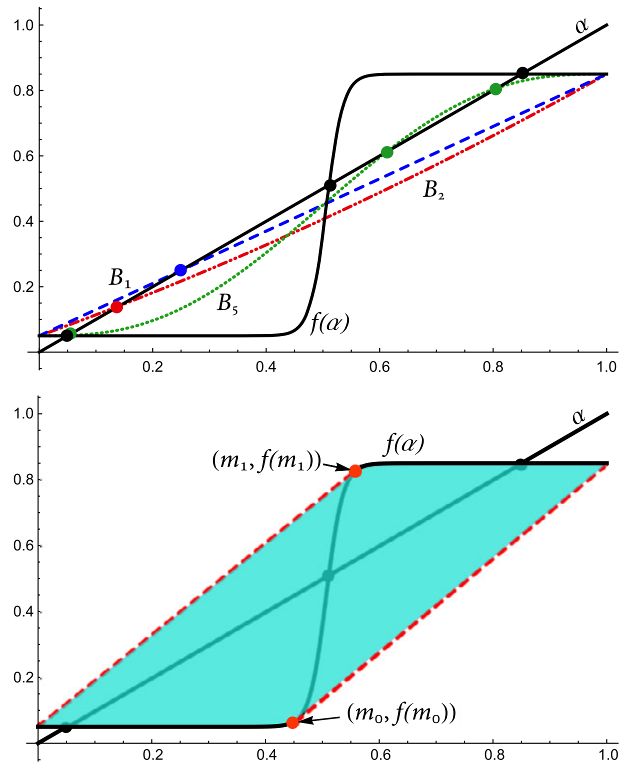

FIGURE 2.—SESIs in a network market: SESI and NE proportions (top) and the convex hull of f (bottom).

the smallest and largest NE proportions. To characterize the location of the SESI proportions relative to the NE proportions more generally, consider the convex hull $\mathrm{co}(f)$ of f depicted in the bottom panel of Figure 2. The lower envelope of $\mathrm{co}(f)$ is f itself from 0 up to the point m_0 and is then the line segment connecting $f(m_0)$ and $f(1)$.[10] Because $f(m_0) < m_0$, the smallest NE proportion is smaller than m_0. Similarly, the upper envelope of $\mathrm{co}(f)$ is the line segment connecting $f(0)$ and $f(m_1)$ up to the point m_1 and is then f itself.[11] Because $f(m_1) > m_1$, the largest NE proportion is larger than m_1. Because $\mathrm{Bern}_k(\alpha; f)$ is the convex combination of the values of f, its graph lies inside $\mathrm{co}(f)$. It is above f between 0 and m_0 and below f between m_1 and 1. Thus, the following holds.

OBSERVATION 3: *If $f(m_0) < m_0$ and $f(m_1) > m_1$, then any SESI proportion is larger than the smallest NE proportion and is smaller than the largest NE proportion.*

6. STATISTICAL INFERENCE AND MONOPOLY THEORY

This section demonstrates how statistical inference can be incorporated into monopoly theory when consumers have a preference for uniqueness, and discusses how the predictions of statistical inference differ from those of rational expectations in this context.

[10]Formally, $m_0 = \arg\max_{\alpha\in[0,1]} \frac{f(1)-f(\alpha)}{1-\alpha}$.
[11]Formally, $m_1 = \arg\max_{\alpha\in[0,1]} \frac{f(\alpha)-f(0)}{\alpha}$.

Consumers have a preference for uniqueness when their consumption utility from a good decreases in the number of individuals consuming the good. Such a preference may arise when the good conveys social status (e.g., jewelry) or when its consumption reflects freedom or independence (e.g., a designer clothing item).

To model preference for uniqueness, suppose an individual has consumption value θ for the good, and a disutility of -1 if he meets one or more individuals who also consume the good in t random encounters. Thus, if an α proportion of the population consumes the good, the expected disutility is $1-(1-\alpha)^t$. For tractability, we will replace $(1-\alpha)^t$ with $e^{-\alpha t}$ and use the consumption utility $u(\theta,\alpha)=\theta-(1-e^{-\alpha t})$ in the analysis.

A monopolist produces the good at zero marginal cost and sets a price p to maximize profit. Consumers observe the price and estimate the expected consumer demand for the good at this price. They then decide whether to purchase the good. The monopolist knows how consumers make decisions and takes this knowledge into account when pricing.

We are interested in comparing the monopolist's profit when consumers use statistical inference and when they use rational expectations. This comparison may be informative about the monopolist's incentives when deciding how much information to release about the expected demand for the good. We first solve an example and then provide a general characterization.

EXAMPLE 5: Suppose that $t=4$ and that the monopolist sets a price p for the good. An individual who purchases the good obtains a utility of $\theta-p-(1-e^{-4\alpha})$. Otherwise, his utility is 0.

Rational expectations require α to be equal to the proportion of consumers with $\theta\geq p+(1-e^{-4\alpha})$, so α has to satisfy $\alpha=1-p-(1-e^{-4\alpha})$. The monopolist's inverse demand is therefore $p=e^{-4\alpha}-\alpha$. Solving the monopolist's profit maximization problem, we obtain that the monopolist's profit under rational expectations is 0.060 with an optimal quantity of 0.135 and an optimal price of 0.448.

Unbiased statistical inference from a single observation implies that when a consumer observes a success, he estimates α to be 1 and hence his utility to be $\theta-p-(1-e^{-4})$. For $p\geq 0.02$, this utility is negative for any $\theta\leq 1$, and thus no consumer will purchase the good in this case. When the consumer observes a failure, he estimates α to be 0 and hence his utility to be $\theta-p$. The proportion of individuals who purchase the good in this case is $1-p$. Thus, at a price $p\geq 0.02$, the SESI demand has to satisfy $\alpha=\alpha\cdot 0+(1-\alpha)(1-p)$ and the inverse demand is $p=1-\frac{\alpha}{1-\alpha}$. Solving the monopolist's profit maximization problem, we obtain that the monopolist's profit is 0.17 with an optimal quantity of 0.29 and an optimal price of 0.59. The monopolist's profit is almost three times larger than with rational expectations.

Unbiased statistical inference from two observations depends on how consumers estimate demand after observing one success and one failure. Suppose that consumers use Beta estimation and so have the estimate $\mathrm{Beta}(1,1)$ in this case. The estimated disutility is then the expected value of $1-e^{-4\alpha}$ with respect to $\mathrm{Beta}(1,1)$, which is about 0.75. The proportion of individuals who purchase the good conditional on this sample is $\max\{0.25-p,0\}$. It is easy to verify that $p\leq 0.25$ is not optimal for the monopolist and thus the monopolist will cater only to consumers who observe two failures. The optimal profit is smaller than with one observation at 0.14, the optimal quantity is also smaller at 0.23, but the optimal price is higher at 0.62.

More generally, we have the following comparative static result.

THEOREM 6: *Fix an unbiased inference procedure that preserves shape and satisfies noise reduction. The monopolist's profit when consumers use statistical inference is larger than when consumers have rational expectations, and it decreases in the sample size.*

PROOF: Suppose the monopolist sets a price p for the good. Fix an inference procedure G and a sample size k. A consumer who observes j successes purchases the good if his θ is larger than $p+1-F_{k,j/k}$, where $F_{k,j/k}$ is the expected value of $e^{-\alpha t}$ with respect to $G_{k,j/k}$. Thus, conditional on observing j successes, the proportion of consumers who purchase the good is $\max\{F_{k,j/k}-p,0\}$. Demand at price p, $\alpha_{k,G}(p)$, is therefore determined by the solution to the equation

$$\alpha=\sum_{j=0}^{k}\binom{k}{j}\alpha^j(1-\alpha)^{k-j}\max\{F_{k,j/k}-p,0\},$$

which can be rewritten as

$$1-\alpha=\sum_{j=0}^{k}\binom{k}{j}\alpha^j(1-\alpha)^{k-j}\min\{1+p-F_{k,j/k},1\}.$$

Because the function $e^{-\alpha t}$ is convex and decreasing, the function F_k is convex and decreasing, and the function $1+p-F_k$ is concave and increasing in α. The function $\min\{1+p-F_k,1\}$ is therefore concave and increasing. We thus have an equation which is similar to Equation (1) and the results of Section 3 continue to hold. That is, for every price p, the demand $\alpha_{k,G}(p)$ decreases in k and is above the rational expectations demand, which is the $\alpha(p)$ that solves $\alpha=1-e^{-\alpha t}-p$. We thus obtain a ranking of the demand functions the monopolist faces, and this ranking implies the ranking of the monopolist's profit. *Q.E.D.*

Thus, the monopolist has no incentive to inform consumers about the demand for the good.

7. STATISTICAL INFERENCE IN TWO-SIDED MARKETS

This section incorporates statistical inference to two-sided markets in which workers and firms need to decide whether to engage in costly search in order to create jobs.

There is a unit mass of workers, and each of them needs to decide whether to search for a job at an idiosyncratic cost θ. Workers' costs are distributed uniformly on $[0,1]$. There is also a unit mass of firms, and each of them needs to decide whether to post a vacancy at an idiosyncratic cost ω. Firms' costs are also distributed uniformly on $[0,1]$. If α workers search for jobs and β firms search for workers, the number of jobs created is given by the matching function

$$m(\alpha,\beta)\equiv\mu\alpha^x\beta^{1-x},$$

where $0<\mu<1$ is the matching friction that prevents full employment even if all workers and all firms participate in the market. As noted by Petrongolo and Pissarides (2001), "the matching function summarizes a trading technology between agents who place advertisements, read newspapers and magazines, go to employment agencies, and mobilize local networks that eventually bring them together into productive matches. The key idea is that this complicated exchange process is summarized by a well-behaved function that gives the number of jobs formed at any moment in time in terms of the number of workers looking for jobs, the number of firms looking for workers, and a small number of other variables.... The stylized fact that emerges from the empirical literature is that there is a stable aggregate matching function of a few variables that satisfies the Cobb–Douglas restrictions with constant returns to scale in vacancies and unemployment."

A worker who searches for a job finds one with probability $\frac{m(\alpha,\beta)}{\alpha}$ when α workers and β firms participate in the market. Assuming that each match creates a surplus of 2 to be split equally between the worker and the firm, the expected utility of a worker who searches for a job is $\frac{m(\alpha,\beta)}{\alpha} - \theta$. Similarly, the expected utility of a firm that posts a vacancy is $\frac{m(\alpha,\beta)}{\beta} - \omega$. Firms and workers who do not participate in the market obtain a utility of 0.

In a NE, all workers and firms form the same belief about the market thickness $(\alpha_{\mathrm{NE}}, \beta_{\mathrm{NE}})$. This belief satisfies the rational expectations assumption whereby if workers and firms best-respond to the belief, actual market thickness is $(\alpha_{\mathrm{NE}}, \beta_{\mathrm{NE}})$. Thus, market thickness on the workers' side α_{NE} has to satisfy that all workers with cost $\theta \leq \frac{m(\alpha_{\mathrm{NE}},\beta_{\mathrm{NE}})}{\alpha_{\mathrm{NE}}} = \mu\alpha_{\mathrm{NE}}^{x-1}\beta_{\mathrm{NE}}^{1-x}$ search for jobs. Applying the same reasoning for firms, we obtain that the two equations that characterize the NE participation are $\alpha = \mu\alpha^{x-1}\beta^{1-x}$ and $\beta = \mu\alpha^{x}\beta^{-x}$. Rearranging, we obtain the following equilibrium equations:

$$\alpha_{\mathrm{NE}}(\beta) \equiv \left(\mu\beta^{1-x}\right)^{\frac{1}{2-x}}, \tag{3}$$

$$\beta_{\mathrm{NE}}(\alpha) \equiv \left(\mu\alpha^{x}\right)^{\frac{1}{1+x}}. \tag{4}$$

We prove below that there is a unique NE with positive participation.

We relax the rational expectations assumption. We postulate that while each worker estimates market thickness on the workers' side accurately, he uses statistical inference to estimate firms' participation. The difference between workers' reasoning ability regarding workers' participation and firms' participation aims to capture situations in which each worker has a good understanding of his side of the market, for example because he observes the decisions of many workers, but he struggles to understand firms' decision making, for example because his access to information on firms' decision making is limited. In order to get a better understanding of firms' decision making, he obtains data on a few firms and uses statistical inference to estimate the market thickness on the firms' side. We make an analogous assumption on firms' reasoning abilities: each firm estimates the market thickness of the firms' side accurately and obtains data on a few workers in order to estimate workers' participation.

DEFINITION: A *SESI in a two-sided market* is a pair $(\alpha_{k,G^w}, \beta_{m,G^f}) \in [0,1] \times [0,1]$ such that:

(i) Proportion α_{k,G^w} of workers search for jobs when each worker obtains k observations about firms' behavior from a Bernoulli distribution with probability of success β_{m,G^f} and best-responds to the estimate he forms according to the inference procedure G^w and α_{k,G^w}, and

(ii) Proportion β_{m,G^f} of firms search for workers when each firm obtains m observations about workers' behavior from a Bernoulli distribution with probability of success α_{k,G^w} and best-responds to the estimate it forms according to the inference procedure G^f and β_{m,G^f}.

To develop the equations that characterize SESIs in this setting, fix a pair of unbiased inference procedures (G^w, G^f) that preserve shape and satisfy noise reduction. Let

$M(\alpha, (k, z))$ denote the expected matching probability of a worker who estimates that α workers are in the market, obtains the sample (k, z) regarding firms' behavior, and uses the inference procedure G^w. Let $M((m, z'), \beta)$ denote the analogous expected matching probability of a firm that observes the sample (m, z') of workers' behavior. Then, the proportion of workers who participate in the market conditional on observing the sample (k, z) is $M(\alpha, (k, z))$ and similarly the proportion of firms is $M((m, z'), \beta)$. The probability that a worker observes the sample (k, z) with $z = j/k$ is $\binom{k}{j}\beta^j(1-\beta)^{k-j}$ and the probability that a firm observes the sample (m, z') with $z' = i/m$ is $\binom{m}{i}\alpha^i(1-\alpha)^{m-i}$. The equations that characterize SESIs are therefore

$$\alpha = \sum_{j=0}^{k} \binom{k}{j} \beta^j (1-\beta)^{k-j} M\big(\alpha, (k, j/k)\big),$$

$$\beta = \sum_{i=0}^{m} \binom{m}{i} \alpha^i (1-\alpha)^{m-i} M\big((m, i/m), \beta\big).$$

The expected matching probability of workers $M(\alpha, (k, z))$ satisfies $M(\alpha, (k, z)) = \mu\alpha^{x-1}M^w(k, z)$, where $M^w(k, z)$ is the expected value of β^{1-x} with respect to the estimate $G^w_{k,z}$. Rewriting the expected matching probability of firms similarly, we obtain the equilibrium equations

$$\alpha_k(\beta) \equiv \left(\mu \sum_{j=0}^{k} \binom{k}{j} \beta^j (1-\beta)^{k-j} M^w(k, j/k)\right)^{\frac{1}{2-x}}, \tag{5}$$

$$\beta_m(\alpha) \equiv \left(\mu \sum_{i=0}^{m} \binom{m}{i} \alpha^i (1-\alpha)^{m-i} M^f(m, i/m)\right)^{\frac{1}{1+x}}. \tag{6}$$

While the environment and the equations characterizing SESIs in two-sided markets are different from those of Sections 2 and 3, we can still use the theory of Bernstein polynomials to characterize SESIs in two-sided markets.

THEOREM 7: *Fix a pair of unbiased inference procedures G^w and G^f that preserve shape and satisfy noise reduction, and sample sizes k for workers and m for firms. Then, there exists a unique SESI with positive employment. In the positive SESI, market thickness on the workers' and the firms' sides, and hence employment, are smaller than in the unique NE with positive employment. Moreover, as k or m increase, market thickness and employment in the positive SESI increase.*

PROOF: The function $\alpha_k(\beta)$ is strictly concave and strictly increasing in β. This is because (1) β^{1-x} is concave and strictly increasing in β, hence (2) $M^w(k, z)$ is concave and strictly increasing in z because G^w is an inference procedure that preserves shape, hence (3) the Bernstein polynomial of $M^w(k, z)$, which is the expression in parentheses in Equation (5), is concave and strictly increasing, hence (4) this expression raised to the power of $\frac{1}{2-x} < 1$ is strictly concave and strictly increasing in β. A similar argument establishes that $\beta_m(\alpha)$ is strictly concave and strictly increasing in α, which implies that its inverse function $\hat{\alpha}_m(\beta)$ is strictly convex and strictly increasing in β. Now, β is part of a SESI if and only if it is a point in which the functions $\alpha_k(\beta)$ and $\hat{\alpha}_m(\beta)$ intersect. Because $\alpha_k(\beta)$

is strictly concave and $\hat{\alpha}_m(\beta)$ is strictly convex and both are increasing, they intersect in at most one positive point. Since the value of both functions at $\beta = 0$ is zero, and at $\beta = 1$, $\alpha_k(\beta) < \hat{\alpha}_m(\beta)$, we obtain that such a point exists. A similar argument can be used to establish that there is a unique NE with positive employment.

Arguments similar to the ones used in Theorems 2 and 3 can be used to show that the function $\alpha_k(\beta)$ lies below the function $\alpha_{\mathrm{NE}}(\beta)$ and that the function $\alpha_{k+1}(\beta)$ lies between these two functions for any $0 < \beta < 1$. Similarly, the function $\hat{\alpha}_m(\beta)$ lies above the function $\hat{\alpha}_{\mathrm{NE}}(\beta)$ and the function $\hat{\alpha}_{m+1}(\beta)$ lies between these two functions. This implies the ranking of the equilibria. *Q.E.D.*

The magnitude of underemployment in a SESI can be substantial. The following example illustrates.

EXAMPLE 6: Suppose $x = 1/2$. Then, the unique positive NE has to solve $\alpha = \mu(\beta/\alpha)^{1/2}$ and $\beta = \mu(\alpha/\beta)^{1/2}$. Participation in this NE is $\alpha = \beta = \mu$ and employment is μ^2. The unique positive SESI with unbiased inference procedures and $k = m = 1$ has to solve $\alpha = \beta\mu(1/\alpha)^{1/2}$ and $\beta = \alpha\mu(1/\beta)^{1/2}$. Participation in the SESI is smaller than in the NE by a factor of μ at $\alpha = \beta = \mu^2$ and employment is smaller by the same factor and is μ^3.

8. CONCLUDING COMMENTS

This paper made modest progress in incorporating statistical inference into games. We postulated that players act as statisticians. They obtain data on the actions of a few other players and use statistical inference procedures such as MLE or Beta estimation to estimate other players' actions. In particular, they do not "think strategically" about how other players make inferences or choose actions. In order to "close" the model, we assumed that the data generating process from which players obtain data is representative of the distribution of players' actions. A solution to this model, which depends on players' sample size and inference procedure, is called a SESI. We developed tools for the analysis of SESIs, and used these tools to study the predictions of SESIs in several settings including competitive and network markets, monopoly pricing, and two-sided markets.

We conclude with three comments. The first is about the interpretation of SESI as the steady state of a dynamic process, the second is about the robustness of the analysis and the results to heterogeneous sample sizes, and the third is about robustness to heterogeneous inference procedures. For simplicity, all comments focus on the setting of Sections 2 and 3 in which the function f is convex, the inference procedures are unbiased, preserve shape, and satisfy noise reduction, and the SESI is unique.

8.1. *The Steady-State Interpretation of SESI*

A SESI is interpreted as a steady state of a dynamic process in which the entry of new statisticians to an existing population does not alter the distribution of actions in the population.

To formalize this interpretation, consider a population of measure 1, and let $0 \leq \alpha_0 \leq 1$ denote the proportion of players in the population who take action A in period 0. In every subsequent period $t = 1, 2, \ldots$, a proportion $1 - \epsilon$ of the population is randomly selected to survive. These players continue to take the same action as in the previous period. The remaining ϵ-proportion leaves the population and is replaced by an ϵ-proportion of new players. The new players obtain data on the previous period's actions, use statistical inference to estimate the proportion of players taking the action, and best-respond to the resulting estimate. Thus, the proportion of players who take action A evolves according to the process

$$\alpha_t = (1-\epsilon)\alpha_{t-1} + \epsilon\bigl(1 - \mathrm{Bern}_k(\alpha_{t-1}; F_k)\bigr) \equiv h(\alpha_{t-1}).$$

A steady state of this process is a proportion α^* that satisfies $\alpha^* = 1 - \mathrm{Bern}_k(\alpha^*; F_k)$, that is, it is a SESI proportion of degree k with respect to the corresponding inference procedure. By the results of Section 3, there is a unique steady state. The following result establishes convergence of the dynamic process to this steady state.

THEOREM 8: *Let $\epsilon < \frac{1}{1+f'(1)}$. Then, for any initial proportion α_0, the proportion α_t of players who take action A converges to the unique SESI proportion as t tends to infinity.*

Theorem 8 continues to hold for a concave function f when $\epsilon < \frac{1}{1+f'(0)}$.[12]

8.2. *Heterogeneous Sample Sizes*

Heterogeneous sample sizes may arise due to differences in individuals' access to information or due to differences in the ability to recall or process information. To model such heterogeneity, let γ_i denote the proportion of individuals who obtain a sample of size i with $1 \le i \le K$ and $\sum_{i=1}^{K} \gamma_i = 1$, and let $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_K)$.

A SESI of degree $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_K)$ with respect to the inference procedure G is a number $\alpha_{\boldsymbol{\gamma},G} \in [0,1]$ such that a proportion $\alpha_{\boldsymbol{\gamma},G}$ of players take action A when, for every $1 \le i \le K$, a proportion γ_i of players obtains i independent observations from a Bernoulli distribution with probability of success $\alpha_{\boldsymbol{\gamma},G}$, and each player best-responds to the estimate he forms using the inference procedure G.

Similarly to Observation 1, any SESI proportion $\alpha_{\boldsymbol{\gamma},G}$ solves

$$1 - \alpha = \sum_{i=1}^{K} \gamma_i \,\mathrm{Bern}_i(\alpha; F_i), \tag{7}$$

where F_i is a function that assigns the value $F_{i,z}$ to every $z \in [0,1]$.

The left-hand side of Equation (7) is identical to the left-hand side of Equation (1). The function on the right-hand side of Equation (7) is a finite convex combination of the functions on the right-hand side of Equation (1). This convex combination inherits the properties of the functions on which it operates that we used in the proofs of Theorems 1 and 2. It therefore follows that (i) for any $\boldsymbol{\gamma}$ and G, there exists a unique SESI, and (ii) the SESI proportion is smaller than the NE proportion.

In order to rank two SESI proportions, $\alpha_{\boldsymbol{\gamma},G}$ and $\alpha_{\hat{\boldsymbol{\gamma}},G}$, we treat $\boldsymbol{\gamma}$ and $\hat{\boldsymbol{\gamma}}$ as discrete distributions over the values $1, \dots, K$. Then, the proportion $\alpha_{\boldsymbol{\gamma},G}$ is strictly smaller than $\alpha_{\hat{\boldsymbol{\gamma}},G}$ if $\boldsymbol{\gamma}$ is first-order stochastically dominated by $\hat{\boldsymbol{\gamma}}$.

To see why, recall that we established in Theorem 3 that $\mathrm{Bern}_i(\alpha; F_i) \ge \mathrm{Bern}_{i+1}(\alpha; F_{i+1})$ with strict inequality for $0 < \alpha < 1$. This implies that for every $0 < \alpha < 1$, the right-hand side of Equation (7) with respect to $\boldsymbol{\gamma}$ is larger than with respect to $\hat{\boldsymbol{\gamma}}$ by the first-order stochastic dominance. Therefore, the SESI proportion $\alpha_{\boldsymbol{\gamma},G}$ is smaller.

[12]The upper bound on ϵ changes because the concavity of f implies that the corresponding Bernstein polynomial of F_k lies below f and its derivative decreases in α.

8.3. *Heterogeneous Inference Procedures*

Heterogeneous inference procedures may arise for various reasons such as a preference for a particular estimation procedure or different levels of confidence in the estimation. To model such heterogeneity, let $G^1, \dots, G^M$ denote the M inference procedures used by players, and let γ_i denote the proportion of players who use the inference procedure G^i.

A SESI of degree k with respect to the distribution $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_M)$ of inference procedures $\boldsymbol{G} = (G^1, \dots, G^M)$ is a number $\alpha_{k,\boldsymbol{\gamma}} \in [0, 1]$ such that a proportion $\alpha_{k,\boldsymbol{\gamma}}$ of players take action A when each player obtains k independent observations from a Bernoulli distribution with probability of success $\alpha_{k,\boldsymbol{\gamma}}$, a proportion γ_i of players form an estimate according to the inference procedure G^i for every $1 \le i \le M$, and each player best-responds to his estimate in choosing an action.

Assume that every G^i is unbiased, preserves shape, and satisfies noise reduction. Fix a convex cost function f and a sample size k, and let F^i_k denote the function that assigns to any $z \in [0, 1]$ the expected value of f with respect to G^i conditional on the sample (k, z).

Similarly to Observation 1, any SESI proportion $\alpha_{k,\boldsymbol{\gamma}}$ has to solve

$$1 - \alpha = \sum_{i=1}^{M} \gamma_i \operatorname{Bern}_k\bigl(\alpha; F^i_k\bigr). \tag{8}$$

Arguments similar to the ones used above and in Section 3 can now be used to establish that (i) for any sample size and any collection of inference procedures, there exists a unique SESI, (ii) the SESI proportion is smaller than the NE proportion, and (iii) the SESI proportion $\alpha_{k,\boldsymbol{\gamma}}$ increases in k.

Thus, the results of Section 3 extend to heterogeneous samples and heterogeneous inference procedures. A challenge for future research is to study settings with dynamic or correlated sampling.

APPENDIX: PROOFS

PROOF OF THEOREM 4. By Jensen's inequality and the convexity of f, the function F_k is larger than the function f. Because G preserves shape and f is strictly convex, the function F_k is strictly convex, so Property 2 implies $\operatorname{Bern}_k(\alpha; F_k)$ is convex and larger than F_k. In fact, $\operatorname{Bern}_k(\alpha; F_k)$ is strictly convex for $k \ge 2$.[13] Similarly to the proof of Theorem 2, one can also show that $\operatorname{Bern}_k(\alpha; F_k) \ge \operatorname{Bern}_{k+1}(\alpha; F_{k+1})$ with strict inequality for $0 < \alpha < 1$ as the relevant arguments rely only on f being convex. Thus, we have the ranking

$$f(\alpha) \le F_k(\alpha) \le \operatorname{Bern}_k(\alpha; F_k) \le \operatorname{Bern}_{k+1}(\alpha; F_{k+1}),$$

where the second and third inequalities are strict for $0 < \alpha < 1$.

Equation (2) implies that the number of NEs is equal to the number of times $f(\alpha)$ intersects the function $1 - \alpha$. Observation 1 implies that the number of SESIs of degree k is equal to the number of times $\operatorname{Bern}_k(\alpha; F_k)$ intersects $1 - \alpha$.

Because $f(0) = F_k(0) = \operatorname{Bern}_k(0; F_k) < 1$ and $f(1) = F_k(1) = \operatorname{Bern}_k(1; F_k) > 0$, the functions $f(\alpha)$ and $\operatorname{Bern}_k(\alpha; F_k)$ intersect $1 - \alpha$ at least once. And because $f(\alpha)$ and $\operatorname{Bern}_k(\alpha; F_k)$ for $k \ge 2$ are strictly convex, they intersect $1 - \alpha$ exactly once. For $k = 1$,

[13]Strict convexity follows because the second derivative of $\operatorname{Bern}_k(x; v)$ for $k \ge 2$ is proportional to a convex combination of expressions of the form $v(\frac{j}{k}) + v(\frac{j+2}{k}) - 2v(\frac{j+1}{k})$ for $0 \le j \le k - 2$.

$\text{Bern}_k(\alpha; F_k)$ is a linear function with a larger slope than that of $1-\alpha$, so they intersect exactly once. This establishes that there is a unique NE and a unique SESI for any sample size.

The ranking of the NE and the SESI proportions of different degrees now follows from the ranking of the functions f, Bern_k, and Bern_{k+1}. *Q.E.D.*

PROOF OF THEOREM 5. By the discussion following the statement of Theorem 5, it suffices to prove the following:

PROPOSITION A1: *For $k \geq 2$, the kth-order Bernstein polynomial of an S-shaped function f is convex if $\delta^k_{k-1} \geq 0$, is concave if $\delta^k_1 \leq 0$, and is otherwise S-shaped. Moreover, if the kth-order Bernstein polynomial of f is convex (concave), then so is the $(k-1)$th-order polynomial of f.*

We first state and prove four observations and then use them to prove the proposition.

Let i be the inflection point of f. For any $\beta > i$ and any $\gamma < \beta$, let $s_\beta(\gamma) = \frac{f(\beta)-f(\gamma)}{\beta-\gamma}$ denote the slope of the line segment $L_f(\gamma, \beta)$ connecting $f(\gamma)$ and $f(\beta)$ on the graph of f. Because f is S-shaped, s_β initially increases and then decreases in γ, implying that $m_\beta = \arg\max_{\gamma<\beta} s_\beta(\gamma)$ is unique. The point m_β is weakly smaller than the inflection point because s_β strictly decreases to the right of the inflection point. If $m_\beta > 0$, then $L_f(m_\beta, \beta)$ is tangent to f at m_β.

OBSERVATION A1: *For any $\zeta \in (m_\beta, \beta)$, the line segment $L_f(\zeta, \beta)$ is strictly below $f(\alpha)$ for $\alpha \in (\zeta, \beta)$.*

PROOF: The line segment $L_f(\zeta, \beta)$ is strictly below the line segment $L_f(\alpha, \beta)$ between α and β because $m_\beta < \zeta < \alpha$. The result follows because $L_f(\alpha, \beta)$ intersects the graph of f at α. *Q.E.D.*

OBSERVATION A2: *If $\beta_2 > \beta_1 > i$, then $m_{\beta_2} \leq m_{\beta_1}$.*

PROOF: By the definition of m_{β_1}, $f'(m_{\beta_1}) \geq s_{\beta_1}(m_{\beta_1})$ with equality if $m_{\beta_1} > 0$. Because $\beta_2 > \beta_1$, $s_{\beta_2}(m_{\beta_1}) < s_{\beta_1}(m_{\beta_1})$ (the slope of $L_f(m_{\beta_1}, \beta_1)$ decreases when it is rotated to $L_f(m_{\beta_1}, \beta_2)$ because $\beta_1 > i$). Thus, $f'(m_{\beta_1}) > s_{\beta_2}(m_{\beta_1})$, implying that $m_{\beta_2} \leq m_{\beta_1}$. *Q.E.D.*

OBSERVATION A3: *If $\delta^k_j \geq 0$, then $\delta^k_h \geq 0$ for $h < j$ with strict inequality for $h < j-1$.*

PROOF: If $\frac{j}{k} \leq i$, then the convexity of f to the left of the inflection point implies the result.

Suppose $\frac{j}{k} > i$. Then, $\frac{j-1}{k} < i$ (otherwise $\delta^k_j < 0$ by the strict concavity of f to the right of i). The strict convexity of f to the left of the inflection point implies the result for $h < j-1$.

To show that $\delta^k_j \geq 0$ implies $\delta^k_{j-1} \geq 0$, assume to the contrary that $\delta^k_{j-1} < 0$. This implies that $f(\frac{j-1}{k})$ is above the line segment $L_f(\frac{j-2}{k}, \frac{j}{k})$. The line segment $L_f(\frac{j-2}{k}, \frac{j}{k})$ has up to two intersection points with f to the left of $\frac{j}{k}$. Let η denote the larger one if more than one exists. Then, $\frac{j-1}{k} > \eta$ (the line segment $L_f(\frac{j-2}{k}, \eta)$ lies above f to the left of η if $\eta > \frac{j-2}{k}$). By construction, $m_{\frac{j}{k}} < \eta$ and thus $m_{\frac{j+1}{k}} < \eta$ by Observation A2. This implies $\delta^k_j < 0$ by Observation A1, which is a contradiction. *Q.E.D.*

OBSERVATION A4: *If* $\delta_{k-1}^k \geq 0$, *then* $\delta_{k-2}^{k-1} \geq 0$. *Similarly*, *if* $\delta_1^k \leq 0$, *then* $\delta_1^{k-1} \leq 0$.

PROOF: Suppose $\delta_{k-1}^k \geq 0$. Then, $\frac{k-2}{k} < m_1$ (otherwise, by Observation A1, $\delta_{k-1}^k < 0$) and the line segment $L_f(\frac{k-2}{k}, \frac{k}{k})$ intersects the graph of f at an intermediate point $\eta \geq \frac{k-1}{k}$ (otherwise, if $\eta < \frac{k-1}{k}$, then $\delta_{k-1}^k < 0$). Because $\frac{k-2}{k} < m_1 < i$ and $\frac{k-3}{k-1} < \frac{k-2}{k}$, the line segment $L_f(\frac{k-3}{k-1}, \frac{k}{k})$ is above $L_f(\frac{k-2}{k}, \frac{k}{k})$ and is therefore above the graph of f to the left of η. Since $\frac{k-2}{k-1} < \frac{k-1}{k} < \eta$, we obtain that $L_f(\frac{k-3}{k-1}, \frac{k}{k})$ is above $f(\frac{k-2}{k-1})$ and the result follows. *Q.E.D.*

PROOF OF PROPOSITION A1. Fix $k \geq 2$. The second derivative of the kth-order Bernstein polynomial of f is

$$\begin{aligned} \mathrm{Bern}_k''(\alpha; f) &= k(k-1)\sum_{j=0}^{k-2}\binom{k-2}{j}\alpha^j(1-\alpha)^{k-2-j}\left(f\left(\frac{j+2}{k}\right) - 2f\left(\frac{j+1}{k}\right) + f\left(\frac{j}{k}\right)\right) \\ &= k(k-1)\sum_{j=0}^{k-2}\binom{k-2}{j}\alpha^j(1-\alpha)^{k-2-j}\delta_{j+1}^k, \end{aligned} \tag{9}$$

that is, it is proportional to a convex combination of $\{\delta_j^k\}_{j=1}^{k-1}$.

If $\delta_{k-1}^k \geq 0$, then all the other δ's are non-negative by Observation A3. The second derivative is therefore non-negative and $\mathrm{Bern}_k(\alpha; f)$ is convex. Similarly, if $\delta_1^k \leq 0$, then so are all the other δ's and $\mathrm{Bern}_k(\alpha; f)$ is concave.

Suppose $\delta_1^k > 0$ and $\delta_{k-1}^k < 0$. Let l be the largest integer such that $\delta_l^k \geq 0$. Then, $l < k-1$. By Observation A3, all the δ_j^k's with $j \leq l$ are non-negative (and at least δ_1^k is positive) and all the δ_j^k's with $j > l$ are negative.

Let $t = \sum_{j=l}^{k-2}\binom{k-2}{j}\alpha^j(1-\alpha)^{(k-2-j)}$. Then t increases in α, and we can write the second derivative as follows:

$$\mathrm{Bern}_k''(\alpha; f) = k(k-1)\left[\frac{P}{1-t}(1-t) + \frac{N}{t}t\right],$$

where $P = \sum_{j=0}^{l-1}\binom{k-2}{j}\alpha^j(1-\alpha)^{(k-2-j)}\delta_{j+1}^k$ is positive and $N = \sum_{j=l}^{k-2}\binom{k-2}{j}\alpha^j(1-\alpha)^{(k-2-j)}\delta_{j+1}^k$ is negative. The second derivative is therefore a convex combination of a positive expression and a negative one. At $\alpha = 0$, $\mathrm{Bern}_k''(\alpha; f)$ is positive, at $\alpha = 1$, it is negative, and it decreases in α. Hence, $\mathrm{Bern}_k''(\alpha; f)$ has a unique root on [0, 1], it is positive to the left of this root, and negative to its right, implying $\mathrm{Bern}_k(\alpha; f)$ is S-shaped.

If $\mathrm{Bern}_k(\alpha; f)$ is convex, then $\delta_{k-1}^k \geq 0$ by the above arguments (otherwise, if $\delta_{k-1}^k < 0$, then $\mathrm{Bern}_k(\alpha; f)$ is either concave or S-shaped). By Observation A4, this implies that $\delta_{k-2}^{k-1} \geq 0$, and thus by the above arguments, $\mathrm{Bern}_{k-1}(\alpha; f)$ is convex. The proof for a concave Bernstein polynomial is analogous. *Q.E.D.*

Q.E.D.

PROOF OF THEOREM 8. By the monotone convergence theorem and the uniqueness of the steady state, it suffices to prove that one of the following two conditions holds for any $t \geq 1$:

(i) $\alpha_{t-1} < \alpha_t < \alpha^*$, or

(ii) $\alpha_{t-1} > \alpha_t > \alpha^*$.

Suppose $\alpha_{t-1} < \alpha^*$. The following four steps establish that condition (i) holds.

STEP 1: $\alpha_{t-1} < \alpha_t$.

This follows from the fact that $1 - \text{Bern}_k(\alpha; F_k) > \alpha$ for any $\alpha < \alpha^*$ and the definition of the dynamic process.

STEP 2: $\text{Bern}'_k(\alpha; F_k) \leq f'(1)$ for any $\alpha \in [0, 1]$ where Bern'_k denotes the derivative of the corresponding Bernstein polynomial.

Because f is convex and the inference procedure is unbiased and preserves shape, the function F_k is convex and lies above f. The Bernstein polynomial of F_k is also convex and lies above F_k. The derivative of the Bernstein polynomial increases in α because of convexity, and its value at 1 is smaller than $f'(1)$ because $\text{Bern}_k(\alpha; F_k) \geq f(\alpha)$ for any α with equality at 1 and both functions are convex.

STEP 3: $h'(\alpha) > 0$ for any $\alpha \in [0, 1]$.

By definition,

$$h'(\alpha) = 1 - \epsilon - \epsilon \,\text{Bern}'_k(\alpha; F_k) \geq 1 - \epsilon - \epsilon f'(1) > 0,$$

where the second inequality is by Step 2 and the third inequality is by the choice of ϵ.

STEP 4: $\alpha_t < \alpha^*$.

By the fundamental theorem of calculus,

$$h(\alpha^*) - h(\alpha_{t-1}) = \int_{\alpha_{t-1}}^{\alpha^*} h'(\alpha)\, d\alpha > 0,$$

where the last inequality is by Step 3. Since $h(\alpha_{t-1}) = \alpha_t$ and $h(\alpha^*) = \alpha^*$, the result follows.

Proving that $\alpha_{t-1} > \alpha^*$ implies condition (ii) is analogous. *Q.E.D.*

REFERENCES

AL-NAJJAR, N. I. (2009): "Decision Makers as Statisticians: Diversity, Ambiguity, and Learning," *Econometrica*, 77 (5), 1371–1401. [1728]

ARTHUR, W. B. (1994): "Inductive Reasoning and Bounded Rationality," *American Economic Review: Papers and Proceedings*, 84 (2), 406–411. [1739]

BATTIGALLI, P., AND D. GUAITOLI (1997): "Conjectural Equilibria and Rationalizability in a Game With Incomplete Information," in *Decisions, Games and Markets*. Springer, 97–124. [1728]

BATTIGALLI, P., S. CERREIA-VIOGLIO, F. MACCHERONI, AND M. MARINACCI (2015): "Self-Confirming Equilibrium and Model Uncertainty," *American Economic Review*, 105 (2), 646–677. [1728]

EBBINGHAUS, H. (1885): *Ueber das Gedächtnis: Untersuchungen zur experimentellen Psychologie*. Duncker & Humblot. [1726]

——— (2003): *Memory: A Contribution to Experimental Psychology*. Genesis Publishing Pvt Ltd. [1726]

ESPONDA, I. (2008): "Behavioral Equilibrium in Economies With Adverse Selection," *American Economic Review*, 98 (4), 1269–1291. [1728]

ESPONDA, I., AND D. POUZO (2016): "Berk–Nash Equilibrium: A Framework for Modeling Agents With Misspecified Models," *Econometrica*, 84 (3), 1093–1130. [1728]

EYSTER, E., AND M. RABIN (2005): "Cursed Equilibrium," *Econometrica*, 73 (5), 1623–1672. [1728]

FUDENBERG, D., AND D. K. LEVINE (1993): "Self-Confirming Equilibrium," *Econometrica*, 61 (3), 523–545. [1728]

HALDANE, J. B. S. (1932): "A Note on Inverse Probability," *Mathematical Proceedings of the Cambridge Philosophical Society*, 28 (1), 55–61. [1730]
JEHIEL, P. (2005): "Analogy-Based Expectation Equilibrium," *Journal of Economic Theory*, 123 (2), 81–104. [1728]
JEWITT, I. (1988): "Justifying the First-Order Approach to Principal-Agent Problems," *Econometrica*, 56 (5), 1177–1190. [1736]
LIANG, A. (2018): "Games of Incomplete Information Played by Statisticians," Working Paper. [1728]
MANTILLA, C., R. SETHI, AND J. C. CÁRDENAS (2020): "Efficiency and Stability of Sampling Equilibrium in Public Goods Games," *Journal of Public Economic Theory* (Forthcoming). [1728]
MILGROM, P. R. (1981): "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12 (2), 380–391. [1729]
NÖLDEKE, G., AND J. PEÑA (2016): "The Symmetric Equilibria of Symmetric Voter Participation Games With Complete Information," *Games and Economic Behavior*, 99, 71–81. [1728]
OSBORNE, M. J., AND A. RUBINSTEIN (1998): "Games With Procedurally Rational Players," *American Economic Review*, 88 (4), 834–847. [1726,1728]
——— (2003): "Sampling Equilibrium, With an Application to Strategic Voting," *Games and Economic Behavior*, 45 (2), 434–441. [1726,1728]
PEÑA, J., L. LEHMANN, AND G. NÖLDEKE (2014): "Gains From Switching and Evolutionary Stability in Multi-Player Matrix Games," *Journal of Theoretical Biology*, 346, 23–33. [1728]
PETRONGOLO, B., AND C. A. PISSARIDES (2001): "Looking Into the Black Box: A Survey of the Matching Function," *Journal of Economic Literature*, 39 (2), 390–431. [1727,1743]
PHILLIPS, G. M. (2003): "Bernstein Polynomials," in *Interpolation and Approximation by Polynomials*. Springer Science & Business Media, 247–290. [1733,1737]
SALANT, Y., AND J. CHERRY (2020): "Supplement to 'Statistical Inference in Games'," *Econometrica Supplemental Material*, 88, https://doi.org/10.3982/ECTA17105. [1729]
SCHACTER, D. L. (1999): "The Seven Sins of Memory: Insights From Psychology and Cognitive Neuroscience," *American Psychologist*, 54 (3), 182–203. [1726]
——— (2002): *The Seven Sins of Memory: How the Mind Forgets and Remembers*. Houghton Mifflin Harcourt. [1726]
SETHI, R. (2000): "Stability of Equilibria in Games With Procedurally Rational Players," *Games and Economic Behavior*, 32 (1), 85–104. [1726,1728]
SPIEGLER, R. (2006a): "Competition Over Agents With Boundedly Rational Expectations," *Theoretical Economics*, 1 (2), 207–231. [1728]
——— (2006b): "The Market for Quacks," *Review of Economic Studies*, 73 (4), 1113–1131. [1728]
——— (2019): "Modeling Players With Random 'Data Access'," Working Paper. [1728]
TVERSKY, A., AND D. KAHNEMAN (1971): "Belief in the Law of Small Numbers," *Psychological Bulletin*, 76 (2), 105–110. [1726]
VUL, E., AND H. PASHLER (2008): "Measuring the Crowd Within: Probabilistic Representations Within Individuals," *Psychological Science*, 19 (7), 645–647. [1726]
ZHU, M., AND A. Y. LU (2004): "The Counter-Intuitive Non-Informative Prior for the Bernoulli Family," *Journal of Statistics Education*, 12 (2), 1–10. [1730]

Co-editor Joel Sobel handled this manuscript.

Manuscript received 26 February, 2019; final version accepted 16 December, 2019; available online 22 January, 2020.