Performance Analysis of Axiomatic Models: A predictive approach to testing for Rationality

María José Boccardi

Brown University *

February 9, 2015

Abstract

The utility maximization assumption, a linchpin of many economic models, is often found to be violated in empirical studies. Revealed preference theory provides a nonparametric test for rationality but it requires only one violation to reject the null hypothesis, leading to spurious rejection of rationality in the presence of agent error. I consider a model of rational choice with error, imposing only the constraints given by GARP and allowing for individual-specific behavior. Using a nonparametric projection technique, I recover the closest rational demand vector to the observed data and I construct the predictive distribution as the expected distribution of choices if these were generated by a utility maximizer agent plus error. Then, I propose a novel set of easures for the suitability of the model based on its ability to deliver accurate predictions, providing a meaningful tradeoff between fit and falsifiability. These measures account for the effect of the number and relaitve distribution of budget sets and their interaction with observed behavior on the amount of information that can be extracted from data to produce informative predictions. The empirical performance of these measures is consistent with the theoretical results and they are shown to outperform popular measures in the literature while exhibiting finite sample and asymptotic desirable properties. I show the extension of the proposed framework to a general class of behavioral models .

1 Introduction

The existence of a utility function that represents preferences is a core assumption in economics, and revealed preference theory provides an elegant axiomatic approach for the necessary and sufficient conditions for such assumption to be valid: the Generalized Axiom of Revealed Preference (GARP henceforth). This axiomatic formulation delivers a sharp test that most data sets violate. In this paper I extend the deterministic axiomatization to allow for a stochastic component gauging the performance of the model by its predictive ability, providing a meaningful trade-off between empirical accuracy and falsifiability.

Addressing the sharpness of the rationality test, the literature has proposed a series of goodness of fit measures for those data sets that do not pass the test, mostly based on an intuitive moment of the data that is related to the adjustments to income needed to remove the violations by making them

*Brown University, Department of Economics, Box B, Providence, RI 029012, e-mail: *Maria_Jose_Boccardi@brown.edu*

no longer feasible. For example[1], Echenique *et al.* (2011) proposes the money pump index defined as the total cost of removing inconsistencies as a percentage of total expenditure. These adjustments have not been justified by a behavioral model that generates the data which encumbers ther interpretation of the statistical significance of the required adjustments. Providing a statistical foundation, Varian (1985) and Fleissig and Whitney (2005) among others[2], propose test statistics to assess the significance of the departures from rationality, where the critical values are bootstrapped from a distribution that bounds above the true distribution under the null of rationality. A problem that plagues all these approaches is the dependency of the results with respect to observed budget sets; since the distribution of economic environments may be such that non or very loose constraints are imposed on data for it to be consistent with GARP. This is known in the literature as the power problem since it relates to the probability of identifying violations to the model when the data was indeed generated by a non rational process.[3] Beatty and Crawford (2011) deals with this problem by proposing a measure of predictive success that discounts from realized fit the size of the target area. This measure does not incorporate information about the effect of observed choices in the ability to identified deviant behavior. Alternatively, Andreoni and Harbaugh (2013) proposes a series of power indices exploiting the seemingly panel structure of data to estimate the power of the rationality test. In this paper I build upon these ideas to propose a combined measure of fit and power, providing a meaningful tradeoff between these, exploiting the information contained in data to construct the most precise forecasts without imposing any further assumptions.

On the other hand, nonparametric econometric approaches have failed to deliver demand estimates that are consistent with rationality unless strong assumptions are imposed regarding the nature of the heterogeneity and preferences, see Lewbel (2001). By construction, I allow for individual specific behavior, provinding not only (set) predictions that are consistent with the model in unseen economic environments, but also predictive distributions that incorporate agent's error.

Gabaix and Laibson (2008) considers predictive precision as one of the seven properties of a good model being desirable

> because they [models with predictive precision] facilitate model evaluation and model testing. (...) A model with predictive precision may even be useful when it is empirically inaccurate.

A good model for data is one that not only allows the researcher to explain observed behavior (fit) but also enables the construction of precise predictions for economic environments that have not yet been observed. In this paper I appraise the performance of the utility maximization model by its ability to deliver precise forecasts given observed data. The framework proposed relies on the assumption that choices are the result of decision maker's maximization of her own utility function, given feasible choices, plus an idiosyncratic error term. This assumption is analogous to Varian (1985) but I impose feasibility

[1] Other notable examples are Varian (1990) and Afriat (1972b)

[2] Echenique *et al.* (2011) also proposes a test statistic based on the money pump index under the assumption that prices are observed with (gaussian) error.

[3] See Andreoni and Harbaugh (2013) for an extensive analysis of the power problem.

constraints explicitly. Feasibility is implicitly prescribed by the nature of observed choices and the model, and imposing it results in demand estimates that could have actually been observed. I recover the closest rational behavior to observed choices in the economic environments on which decisions were made, by applying the nonparametric projection technique proposed by Kocoska (2012). Then, I construct the predictive distribution as the expected distribution of choices given the model and the recovered error process. Finally, I assess the performance of the model by sizing the tightness of the predictive distribution.

The predictive distribution is constructed based on the information that can be inferred from data by imposing the model. First, I project the data onto the set of rational choices. The projection technique adopted identifies, from the economic environments in which decisions were made, the rational vector that is the closest to the data. Then, I define the predicted set as the subset of feasible choices that would be consistent with the recovered demand if they were to be observed in the budget set for which the forecast is constructed. Finally, the predictive distribution is defined as the distribution of expected choices under the assumption that choices are given by a rational component and an additive error process. The distribution of the former is given by the distribution of choices over the predicted set, while the distribution for the latter is estimated from observed data. I measure the predictive accuracy of the model by: (i) the size of the shortest $\alpha-$ level credible set for predicted choices to be consistent with the model and (ii) information content of the predictive distribution. These measures can be constructed for all data sets, those that satisfy rationality or not, which permits not only the construction of consistent predictions even when the subjects are observed to fail rationality, but also the comparison of the performance of the model among rational subjects and of these with non rational ones.

The proposed measures based on the predictive distribution provide a meaningful trade off for empirical accuracy (fit) and falsifiability (power); as lower is the fit, higher the variance and therefore lower the informational content of the predictive distribution. Additionally, as more observations are revealed from demanding economic environments, the predicted set shrinks, reducing the entropy of the predictive distribution, i.e. the uncertainty to predict outcomes from the predictive distribution. Finally, I show that the proposed framework can be extended to a general class of behavioral models and be used as a tool for model comparison.

The empirical performance of the predictive measures is studied in its application to the experimental data set from Choi *et al.* (2007a) and Monte Carlo simulations. I show that the predictive ability measures reflect fit and power and that, conditional on fit, standard goodness of fit measures proposed in the literature have no, or slightly negative, effect on the capacity of the model to generate informative predictions given observed data. The simulations studies show that the predictive accuracy deteriorates as noisier is the data and improves when more observations are considered due to the shrinkage of the predicted set.

This paper advances the literature in several directions. First, by extending the model allowing for a idiosyncratic error process I provide a statistical and behavioral model for the assessment of fit and the construction of the predictive distribution; even for those data sets that do not meet GARP constraints. Second, by considering the predictive accuracy of the model, the proposed measures account for the goodness of fit and power problem; therefore are correlated to other measures of fit proposed in the literature but provide further information in terms of power. Also, considering the predictive accuracy provides a meaningful trade off for fit and power in terms of the ability to generate precise predictions. Third, the proposed measures account for the power of the test as in Beatty and Crawford (2011) but also exploits the interaction between observed choices and the relative distribution of budget sets similar to Andreoni and Harbaugh (2013), therefore does not provide the same ordinal results as other measures that combine fit and power. Fourth, I explicitly account for feasibility constraints which affects the distribution of the effective error process and evidence the endogeneity of the error process. In doing so, I formally define the DGP which provides the basis for inference and show that, asymptotically, the predictive distribution converges to the distribution of the effective error process -that accounts for feasibility constraints-; therefore the distribution of a suitable test statistic can be bootstrapped accordingly. Fifth, I present a series of computationally feasible algorithms to recover the distribution of the underlying error process even in small samples. Sixth, I present the extension of the proposed methodology for a general class of axiomatic models that allows for the comparison across data sets and across nested and non-nested models of economic behavior.

The outline of the paper is as follows. Section 2 presents the main results of revealed preference theory and its testable implications. Section 3 provides the theoretical framework. First, I present the extension of the model to allow for a stochastic component and its implications in terms of identification. Second, I present the projection technique. Then, I construct the predictive distribution and predictive accuracy measures. Finally I discussed their asymptotic and finite sample properties. Section 4 shows the comparison of the proposed measures with existing measures in the literature. Section 5 shows the empirical performance in the experimental data from Choi *et al.* (2007a) and Monte Carlo simulations. In section 6 I extend of the proposed framework to general models of economic behavior. Section 8 offers a review of the literature concerning rationality testing and how the approach followed in this paper compares to it. Finally, section 9 concludes.

2 Rationality model

Revealed preference theory provides a nonparametric condition on consumer's choices that is necessary and sufficient for observed behavior to be consistent with utility maximization: GARP, connection that follows from Afriat (1967). The axiomatization provided by GARP is appealing since it does not rely on any assumption on the functional form for the utility function and does not require any homogeneity assumption on preferences across individuals; therefore it can be applied to individual's data even in small samples.

Theorem 1 (Afriat's theorem) *Given data for choices and prices (p^j, x^j) for $j = 1, \ldots, J$ the following conditions are equivalent*

1. *There exists a non-satiated utility function $u(x)$ that rationalizes the data, that is,*

$$\forall j \quad u\left(x^j\right) \geq u(x) \;\; \forall x \quad \textit{such that } p^j x^j \geq p^j x$$

2. *The data satisfies GARP*

3. *There exists a positive solution $\left(u^j, \lambda^j\right)$ to the set of linear inequalities*

$$u^j \leq u^i + \lambda^i p^i \left(x^j - x^i\right) \qquad \forall i, j = 1, \ldots, J$$

4. *There exists a non-satiated, continuous, monotone and concave utility function $u(x)$ that rationalizes the data*

Proof. Fostel *et al.* (2004) ■

The concept of (directly) revealed preference theory was first introduced by Samuelson (1938), Samuelson (1948). Basically, an alternative being chosen from a set reveals information about decision maker's own preference relation, the chosen alternative is revealed to be better to the non-chosen ones. Houthakker (1950) extends the work by Samuelson by imposing transitivity on the direct revealed preferred relation. Formally,

Definition 1 (Directly Revealed Preferred) *x^j is directly revealed preferred to x if $p_j^T x_j \geq p_j^T x$, and it is strictly revealed preferred if $p_j^T x_j > p_j^T x$*

Definition 2 (Revealed Preferred) *x_j is revealed preferred to x if there is a chain of directly revealed preferred bundles linking x_j to x*

The generalized axiom of revealed preference (GARP) proposed by Varian (1982) is equivalent to the "cyclical condition" stated in the original version of Afriat's theorem. Formally,

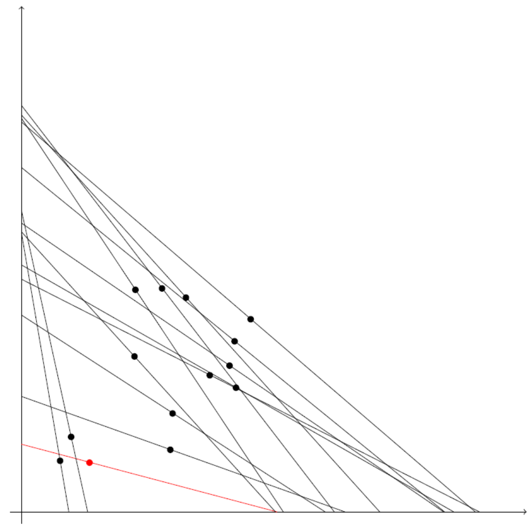
Definition 3 (Generalized Axiom of Revealed Preference (GARP)) *If x_j is revealed preferred to x, then x is not strictly revealed preferred to x_j*

From definitions 1 and 2, GARP can be restated as follows,

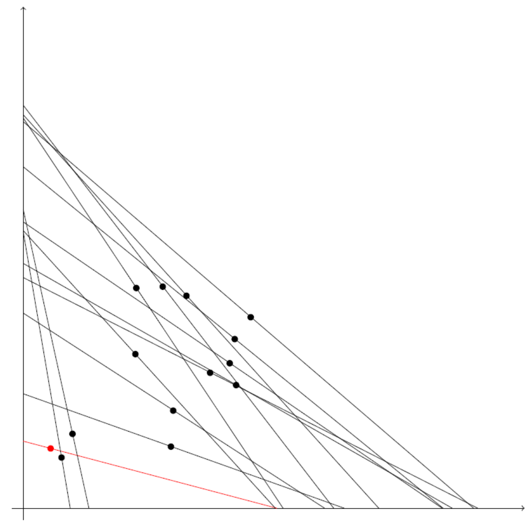
Definition 4 (GARP - A) *Given a data set $\left\{(p^j, x^j)\right\}_{j \in J}$ define the matrix $A_{J \times J}$ as the revealed preference matrix, where $a_{jk} \equiv \langle p^k, x^j - x^k \rangle$. A data set satisfies GARP if for every chain $\{i, j, k, \ldots, r\} \subset J$, $a_{ij} \leq 0, a_{jk} \leq 0, \ldots, a_{ri} \leq 0$[4] implies that all terms are zero.*

A necessary and almost sufficient condition for GARP is given by the following proposition,

[4] Note that $a_{jk} \leq 0$ ($a_{jk} < 0$) implies that x_j is (strictly) directly revealed preferred to x_k



(a) Data consistent with GARP



(b) Data inconsistent with GARP

Figure 1: Sharpness of GARP test. One violation and the data would be declared as inconsistent

Proposition 1 *If GARP is satisfied then there exists a permutation A' of A such that $a'_{kj} \geq 0$ for all $j < k \leq J$.*

Proof. See Appendix A. ■

The above condition only provides a necessary condition but, it turns out, that a sufficient condition can be obtained by strengthening the inequality condition.

Proposition 2 (Sufficient condition) *If there exists a matrix A' obtained from the revealed preference matrix A defined by $a_{jk} = \langle p_j, x_k - x_j \rangle$ by symmetric column/row swaps such that $a'_{jk} > 0$ for all $k < j \leq J$ then GARP is satisfied.*

Proof. See Appendix A. ■

Proposition 3 *The condition of proposition 1 is necessary but not sufficient. The condition of proposition 2 is sufficient but not necessary.*

Proof. See Appendix A. ■

Rationality tests based on GARP deliver a sharp pass/fail result, where in real life most data sets violate the deterministic axiomatization. One violation to rationality and the data would be declared as inconsistent. Consider figure 1, both panels consist in 14 observations, the data in panel (a) is consistent with GARP, while the data in panel (b) is not, even when only one of the observations differs from panel (a) to panel (b). The literature has dealt with this problem by proposing measures of fit based on the extent of the income adjustment to budget sets required to remove inconsistencies by making them no longer feasible. In a more natural exercise, one may think on the minimal perturbation to data required for perturbed choices to be consistent with GARP in the economic environments in which decision maker chooses. In particular, consider the problem of finding a perturbation ε^j to choices x^j such that $x^j + \varepsilon^j$ satisfies proposition 1 and remains feasible, i.e., $x^j + \varepsilon^j \geq 0$ and $(p^j)^T \left(x^j + \varepsilon^j\right) = (p^j)^T x^j$, that minimizes some positive function of the norm of ε. In the following section I present the statistical framework that extends GARP to allow for a stochastic component and the projection technique that permits the recovery of such perturbation.

3 Framework

3.1 Set up

The axiomatization provided by GARP delivers a deterministic test for the utility maximization model. In order to understand the statistical validity of the model, the economic model should be embedded in a statistical model. Therefore, I expand the axiomatic model by incorporating an additive stochastic component to observed choices. Hence, the data is assumed to be generated as the result of decision maker's maximization of her own utility function and an additive error process. This assumption is consistent with the aim of this paper[5], i.e. judge the empirical performance of the utility maximization model by its predictive ability. In section 6 I extend the methodology proposed in this paper for the comparison of a broader class of behavioral model. I impose that observed and rational choices are feasible given observed prices, therefore I do not consider stochasticity on prices, but the framework can be adapted to this case. The chosen specification is consistent with GARP without adding further structure and provides a natural definition for prediction.[6]

Formally, the econometrician observes choices made by an individual that faces J decision problems. The observed data consists of choices $\left\{x^j\right\}_{j=1}^J$ and prices $\left\{p^j\right\}_{j=1}^J$ with $x^j \in X \subseteq \mathbb{R}_+^L$ and $p^j \in P \subseteq \mathbb{R}_{++}^L$ for all $j = 1, \ldots, J$. Assuming local non satiation[7], observed prices define uniquely, up to a proportionality factor, the economic environment realized. Let $B : P \times X \to X$ be the feasible consumption bundle correspondence (set of choices that satisfy the budget constraint) defined by imposing the budget and non-negativity constraints, that is

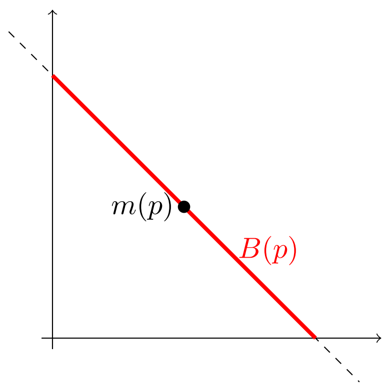
$$B\left(p^j, x^j\right) \equiv \{x \in \mathbb{R}_+^L : (p^j)^T x \leq (p^j)^T x^j\}$$

I adopt the following notation: $y \equiv \left(y^1, \ldots, y^J\right)$ and $\mathbf{Y}(\cdot) \equiv \prod_{j=1}^J Y(\cdot)$. Then $\mathbf{B}(p, x) = \prod_{j=1}^J B(p^j, x^j)$ is the set of feasible consumption bundles vectors.

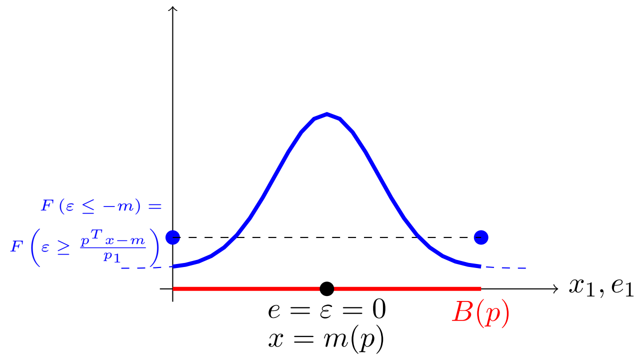
[5] Alternative specifications for the error process that add further structure have consequences about the testing exercise considered. For example, if I were to allow for stochasticity on preference orders it would imply that the model in mind is some variation of RUM.

[6] If we further impose smooth assumptions on demand (notice that this is stronger than continuous differentiability on the utility function) there correspondent RUM approach.

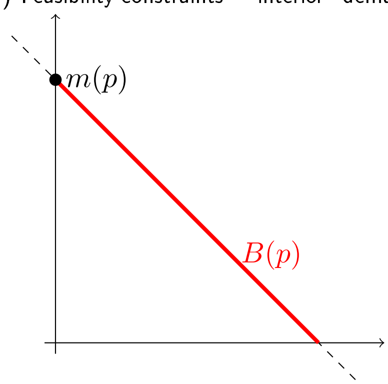
[7] **Definition (Local Non Satiation)** For any $x \in X$ and every $\varepsilon > 0$ there exists a $y \in X$ such that $\|x - y\| \leq \varepsilon$ such that y is preferred to x.



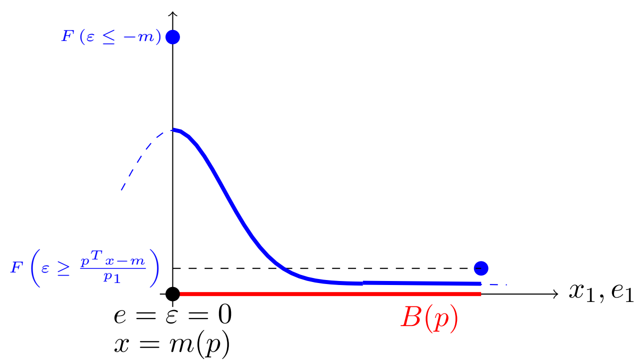
(a) Feasibility constraints - "interior" demand



(b) Distribution of errors/observed choices - "interior" demand



(c) Feasibility constraints - "corner" demand



(d) Distribution of errors/observed choices - "corner" demand

Figure 2: Distributional assumptions: The latent error process is assumed to be independent of the systematic demand process but, the imposition of feasibility constraints on observed demand result on a truncation in the effective distribution of the error process that depends on underlying demand.

The model is assumed to be a correspondence $\mathbf{M} : \mathbf{B}(p, x) \to \mathbf{B}(p, x)$ that maps from the set of feasible choices to the subset of feasible choices that is consistent with GARP, that is

$$\mathbf{M}(\mathbf{B}(p, x)) \equiv \{m \in \mathbf{B}(p, x) : m \text{ satisfies GARP}\}$$

Latent choices are generated by a rational component, $m \in \mathbf{M}(\mathbf{B}(p, x))$, and an additive idiosyncratic component, $\varepsilon \in \mathbb{R}^{L \times J}$, orthogonal to the model, i.e. $\varepsilon \amalg \mathbf{M}(\mathbf{B}(p, x))$. Notice that, by definition, observed and rational choices are feasible, i.e. $x \in \mathbf{B}(p, x)$, and $m(p) \in \mathbf{M}(\mathbf{B}(p, x)) \subseteq \mathbf{B}(p, x)$. These two conditions imply that the domain for the effective error process, e, is restricted to the set such that perturbed choices, $m + e$, are feasible. Formally, let e be the effective error process then,

$$dom(e|m(p)) \equiv \left\{e \in \mathbb{R}^{L \times J} : (m(p) + e) \in \mathbf{B}(p, x) \text{ and } m(p) \in \mathbf{M}(\mathbf{B}(p, x))\right\} \equiv \prod_{j=1}^{J} \left(B\left(p^j, x^j\right) - m_j(p)\right)$$

The feasibility condition imposes two set of constraints: affordability/income neutrality and nonnegativity constraints. First, the perturbed demand should be affordable. By local non satiation, imposing affordability implies that the effective error process should be income neutral, i.e. $p'e \equiv 0$, that is, distribution of the error process is constrained to the hyperplane orthogonal to the price vector. On the other hand, nonnegativity constraints truncate the distribution of the error process to be such that the observed demand lies in the positive orthant. I assume that, the truncation operates as a standard censored model, where the distribution of the error process is given by the latent process in the interior of the feasible set, and mass probability in the closure of the feasible set. This approach is consistent with the utility maximization hypothesis, by continuity guaranteed by Afriat's Theorem, **?**. These two set of constraints imply that the effective error process cannot be assumed to be independent from the underlying rational demand, since, because of the feasibility constraints, $m(p)$ defines the truncation of the error process. Figure 2 shows the effect of the underlying error process on the effective distribution of the error process.

Formally, consider the following assumptions.

Assumption 1 (DGP) *Data is generated by a latent process given by*

$$x^*(p) = m(p) + \varepsilon$$

where $m(p) \in \mathbf{M}\left(\mathbf{B}(p,x)\right)$ *and* $p'\varepsilon = 0$. *Then, observed data is given by*

$$x(p) = \begin{cases} x^*(p) & \textit{if } x^*(p) \geq 0 \\ argmin_{a \in B(p,x)} \|x^*(p) - a\| & \textit{if } \exists j \in \{1, \ldots, J\} \textit{ and } l \in \{1, \ldots, L\} \textit{ s.t. } x_l^{*j} < 0 \end{cases} \tag{1}$$

where $x(p) \in \mathbf{B}(p,x)$ *and* $e \sim F_{\varepsilon|m(p),p'\varepsilon=0}$ *is the effective error process given equation (1), i.e*

$$e \textit{ is such that } x(p) = \tilde{m}(p) + e \textit{ for some } \tilde{m}(p) \in \mathbf{M}\left(\mathbf{B}(p)\right)$$

Assumption 2 (Unconstrained Error Process) *The unconstrained process* $\varepsilon \in \mathbb{R}^L$ *is assumed to be i.i.d. with continuous and symmetric p.d.f. such that* $E(\varepsilon) = F_\varepsilon^{-1}\left(\frac{1}{2}\right) = 0$, $\frac{\partial f(\varepsilon)}{\partial \varepsilon}\varepsilon_j < 0$ *for all* $\varepsilon_j \neq 0$ *and* $\varepsilon \amalg m$, *for some* $m \in \mathbf{M}\left(\mathbf{M}(p,x)\right)$.

Assumption 1 and 2 jointly define the distribution of the observed error process, $F_e = F_{\varepsilon|m(p),p'\varepsilon=0}$ that is given by the constraint of F_ε to the set of affordability and nonnegativity constraints, i.e.,

$$f_e = f_{\varepsilon|m(p),p'\varepsilon=0} = \begin{cases} f_{\varepsilon|p\varepsilon=0} & \text{if } \varepsilon \in int\left(A_\varepsilon\right) \\ \int_a dF_{\varepsilon|p\varepsilon=0} & \text{if } \varepsilon \in cl\left(A_\varepsilon\right) \text{ and } a = argmin_{x \in A_\varepsilon} \|a - x\| \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $A_{\varepsilon,m(p)} = \left\{\varepsilon \in \mathbb{R}^{L\times J} | m(p) + \varepsilon \in \mathbb{R}_+^{L\times J} \text{ and } p'\varepsilon = 0\right\}$

This approach is agnostic in terms of the (rational) theory of underlying behavior, imposing just the minimal set of restrictions on choices that ensures a rationalizable data set. The additive structure of the error process resembles the assumption made in Varian (1985), Fleissig and Whitney (2005), Hjertstrand (2013), Blundell *et al.* (2008) among others; but I further impose feasibility conditions, that, as discussed above, naturally follow from the definition of the model. These constraints imply that underlying behavior is feasible and, therefore, consistent with the model; but induce dependecy between the error process and systematic component.

3.2 Identification of rational demands

The proposed framework remains agnostic about the specific demand process that generated the data, identifying the set of demand vectors that are consistent with the model. For any sample of size J, the constraints imposed by GARP identify the set of demand vectors that can be rationalized as the utility maximizers for some well-behaved utility function, as follows from theorem 1. For any finite sample, $\mathbf{M}(\mathbf{B}(p))$ is a correspondence and not a function. If the data is consistent with GARP, as new information arrives from budget sets that impose binding constraints in data, the set of all possible well-behaved preference relations that are consistent with the data shrinks.

Assumption 3 (Density data) *Let $\tilde{B}^n \equiv \left\{(x^i, p^i)\right\}_{i=1}^n \in X \times P$ be an increasing sequence of sets, i.e. $\left(\tilde{B}^n \subset \tilde{B}^{n+1} \subset \dots\right)$ such that $\lim_{n\to\infty} \bigcup_n \tilde{B}^n$ is dense in $X \times P$.*

Theorem 2 (Preference - Convergence) *Let $m\left(\tilde{B}^n\right) \in \mathbf{M}\left(\mathbf{B}(p)\right)$ be a consistent demand vector given data $\left\{x^i, p^i\right\}_{j=1}^n$. Let $\mathcal{R}_{\{x^i,p^i\}_{j=1}^n}$ be the set of $m\left(\tilde{B}^n\right)$-equivalent preference orders, i.e. the set of preference orders that rationalizes $m\left(\tilde{B}^n\right)$. Let $\tilde{B}^n \equiv \left\{(x^i, p^i)\right\}_{i=1}^n \in X \times P$ be an increasing sequence of sets, i.e. $\left(\tilde{B}^n \subset \tilde{B}^{n+1} \subset \dots\right)$ then $\mathcal{R}_{\{x^i,p^i\}_{j=1}^n} \supseteq \mathcal{R}_{\{x^i,p^i\}_{j=1}^{n+1}} \supseteq \dots$.*

Proof. See Appendix B ■

In finite samples, non-uniqueness is unavoidable, but the set of consistent underlying preferences shrinks as the sample size increases. Mas-Colell (1978) shows that under a boundary condition, if the preference generating the data are strictly convex and monotone, and demands income Lipschitzian, the sequence of preferences constructed from a finite sample by imposing GARP have a unique limit as data gets dense.

Theorem 3 (Mas-Colell (1978)) *Let $m\left(\tilde{B}^n\right)$ be consistent with SARP and a boundary condition*[8] *Let $\tilde{B}^n \equiv \left\{(x^i, p^i)\right\}_{i=1}^n \in X \times P$ be an increasing sequence of sets, i.e. $\left(\tilde{B}^n \subset \tilde{B}^{n+1} \subset \dots\right)$ such that satisfies assumption 3. Suppose that for every n, $\mathcal{R}_{\{x^i,p^i\}_{j=1}^n}$ is a continuous, convex, monotone preference relation with the property that for every $(x,p) \in \tilde{B}^n$, $m\left(\tilde{B}^n\right) \subset$*

[8]**Definition (SARP)**If x_j is revealed preferred to x, and $x_j \neq x$, then x is not revealed preferred to x_j

Definition (Boundary Condition) A demand function h satisfies the boundary condition if $\{x^n, p^n\} \to \{x, p\} \notin \mathbf{B}(p, x), p^T x > 0, \{x^n, p^n\} \in \mathbf{B}(p, x)$ implies that $\|h(x^n, p^n)\| \to \infty$

$\left\{y \in \mathbf{B}(p,x) : p^T y \leq p^T x \text{ and if } p^T v \leq p^T x \Rightarrow y \mathcal{R}_{\{x^i,p^i\}_{j=1}^n} v\right\}$ *then if demand is income Lipschitzian,* $\mathcal{R}_{\{x^i,p^i\}_{j=1}^n} \to \mathcal{R}$ *where* $\mathcal{R}$ *is the unique, continuous, monotone, strictly convex preference relation generating the demand.*

Proof. Refer to Mas-Colell (1978). ■

Given a consistent demand vector, the class of equivalent preference relations that rationalizes demand is well defined by GARP. Under conditions on the relative distribution of budget sets and some regularity conditions it is expected that this sets narrows down. Moreover, the set of vectors that, given observed data, are consistent with the model i.e. $\mathbf{M}(\mathbf{B}(p,x))$ shrinks as the number of observations increases, provided that the newly observed budget sets impose binding constraints in data to satisfy the model.

Theorem 4 (Convergence of $\mathbf{M}(\mathbf{B}(p,x))$) *Let* $\tilde{B}^n$ *be an increasing sequence of sets that satisfies assumption 3. Then, for any* $n \in \mathbb{N}$ *there exists a* $m > n$ *and a* $x^n \in \mathbf{M}(\tilde{B}^n)$ *such that* $\{x^n, y\} \notin \mathbf{M}(\tilde{B}^n \cup \{x^m, p^m\})$.

Proof. See Appendix B ■

3.3 Projection technique

Extending the model to allow for an error process implies that it needs not to be the case that observed data is consistent with GARP. As it was established in the previous section, relying solely on the restrictions imposed by GARP delivers bounds for the estimation of the underlying demand. Blundell *et al.* (2008) discussed the nature of this problem and proposed a semiparametric estimator for demands based on a minimum distance criterion, and, relying on the partial identification literature, showed the relevance of such distance to inference. Blundell *et al.* (2014) identifies nonparametric bounds for demand constructed from revealed preference restrictions and shows their asymptotic properties.

Relying uniquely on the constraints proposed by GARP, I propose to recover a rational demand process by projecting observed data onto the set of consistent demand vectors. In particular, define the projection as the exercise of recovering the vector of consistent choices that is the closest to the observed data, as displayed in the figure 3.

Definition 5 (Projection of observed choices onto the model) *The problem is to find the projection* $\widehat{x}(p)$ *of* $x(p)$ *onto* $\mathbf{M}(\mathbf{B}(p))$ *as the solution of*

$$\begin{aligned} \widehat{x}(p) \quad &\equiv \quad arg \inf_m \sum_{j=1}^{J} \|x^j(p) - m^j\| \\ s.t \quad & m \in \mathbf{M}(\mathbf{B}(p)) \end{aligned}$$

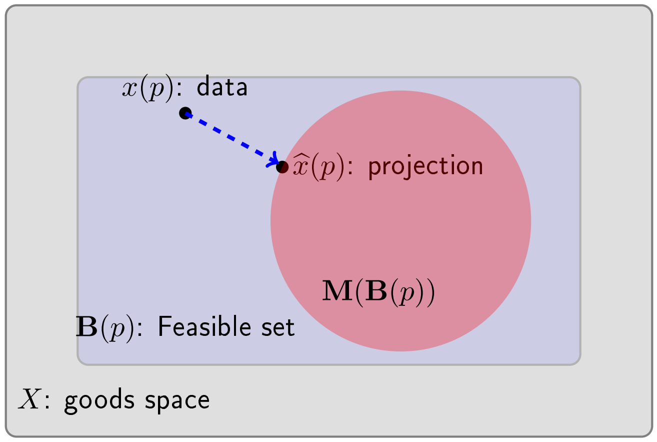


Figure 3: Projection: $\widehat{x}$ is defined as the closest vector to the data that is consistent with GARP

The projection is performed following Kocoska (2012). The author proposes a series of derivative-free algorithms to minimize the perturbation needed to make the perturbed demand consistent with proposition 1. Appendix C provides a theoretical foundation and details for such optimization process.

One may argue that distance in the consumption space is not the relevant metric. The modeler may care about some policy implications and therefore the utility or the budget space may be the relevant one; or may consider that errors in the consumption space may not seem relevant for consumers as the monetary or utilitarian cost of them. Let $\rho : \mathbb{R}^L \to \mathbb{R}^M$ be any positive monotone transformation of the absolute value of the error process that reflects the relevant metric for the modeler, the employed technique is flexible to any such ρ.

3.4 Predictive distribution

A good model does not only allow the researcher to explain observed behavior but enables the construction of informative predictions in economic environments that have not yet been observed. Gabaix and Laibson (2008) recognizes that

> "predictive precision is infrequently emphasized in economics research (...) Models that make weak predictions (or no predictions) are limited in their ability to advance economic understanding of the world (...) Models with predictive precision are easy to empirically test and when such models are approximately empirical accurate they are likely to be useful".

I consider the predictive accuracy of the model to assess the performance of the model measured by the informational content of the predictive distribution. By construction it also provides a meaningful trade off between empirical congruence and falsifiability, in the understanding than even when empirical inaccurate, models that deliver strong predictions are desirable.

The predictive distribution is constructed as the expected distribution of choices on a new economic environment, given assumptions 1, 2 and observed data. First, I construct the set of alternatives that are consistent with rationality given projected data, called supporting set. For each alternative in the supporting set, s, I compute the expected distribution of observed choices conditional on s being the rational underlying demand. This distribution is constructed by first estimating the distribution of ε from the projection residuals, imposing assumptions 1 and 2; and then, imposing the structure given by equation 2 in the new economic environment, centered in s. Finally, I construct the unconditional distribution of choices by integrating the above distributions with respect to some prior distribution over choices conditional on choices lying on the supporting set. This distribution is provided by the econometrician and it is interpreted as the assumed distribution of choices in the space of alternatives if no further information is provided. By default, I consider a uniform distribution, that can be conceptualized as an uninformative prior in the Bayesian sense, and consistent with Becker's definition of irrational behavior Becker (1962).

Given a rational demand, GARP restrictions prescribe the set of alternatives that are consistent with the model in the new economic environment, set that is known as the Varian's supporting set, Varian (1982). The projection exercise delivers a demand estimate that is consistent with the model based on observed data, that can be used to construct the Varian supporting set in a new economic environment.

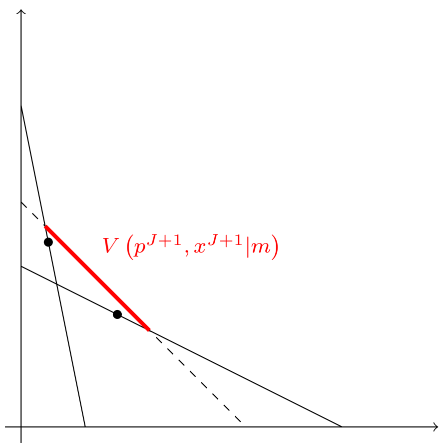
Definition 6 (Varian's supporting set) *Let $m \equiv \left(m^1, \ldots, m^J\right)$ be a vector of choices consistent with rationality, that is, $m \in \mathbf{M}(\mathbf{B}(p, x))$. Given a new economic environment $B\left(p^0, x^0\right)$, the Varian's supporting set is defined as*

$$V\left(p^0, x^0 | m\right) = \left\{x \in B(p^0, x^0) |\, (m, x) \in \mathbf{M}\left(\mathbf{B}\left(\{(p, p^0), (m, x^0)\}\right)\right)\right\} \tag{3}$$

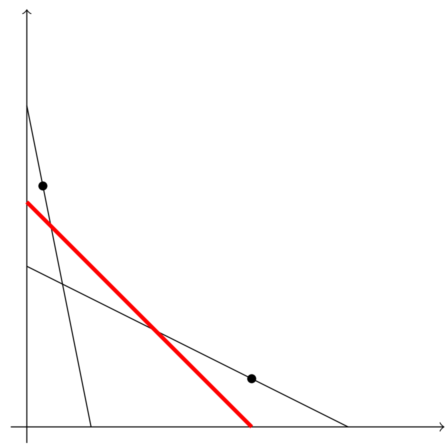
Consider figure 4a. Data is given by the two solid lines and the dots as display in the figure and consider the case of predicting choices in the dashed budget set. In this new budget set, the set of feasible alternatives (assuming local non satiation) that are consistent with the previously observed choices are the ones lying on the bold segment. The relative size of the predicted area depends on the number of observations, the relative distribution of observed and new budget sets and consistent demand vector used for the projection, as shown in panels 4a-4b and 4c-4d. In particular, the supporting set is expected to be smaller as bigger (and more demanding) the data set is.

Proposition 4 (Properties Supporting set) *Let $\{x^0, p^0\}$ be the economic environment for which the supporting set is constructed. Let $\tilde{B}^n$ be an increasing sequence of budget sets as in assumption 3, and let $m^n \in \tilde{B}^n$ then*

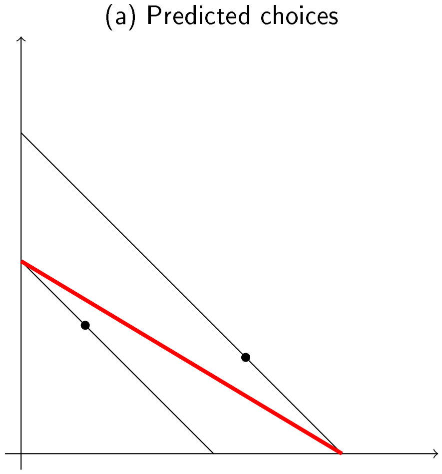
1. *$V\left(p^0, x^0 | m^n\right) \neq \emptyset$ if and only if $m^n \in \mathbf{M}(\mathbf{B}(\tilde{B}^n))$*
2. *$V\left(p^0, x^0 | m^n\right)$ is a convex set*
3. *There exists some $k > n$ such that $V\left(p^0, x^0 | m^k\right) \subset V\left(p^0, x^0 | m^n\right)$*



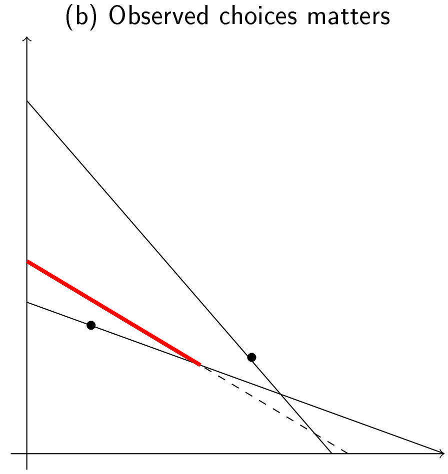
(a) Predicted choices



(b) Observed choices matters



(c) New $\mathbf{B}(\mathbf{p}, \mathbf{x})$ matters



(d) Observed $\mathbf{B}(\mathbf{p}, \mathbf{x})$ matters

Figure 4: Forecasting consumer's behavior

Proof. Follows from theorem 4 and Blundell *et al.* (2008). ■

Behavior in a new economic environment is expected to be consistent with the assumed DGP process, i.e. given a rational component plus an idiosyncratic error process. The distribution of this error process can be estimated from the projection residual, provided that the structure implied by equation 2 is incorporated. The projection error is defined as the residuals from the projection, i.e.

$$\widehat{e} = x(p, x) - \widehat{x} \tag{4}$$

where $\widehat{x}$ solves the problem defined in 5. Given a parametric assumption on F_ε, the relevant parameters can be estimated by maximum likelihood from equation 2; for small samples I provide a more accurate method to recover the parameters in section 3.6.

Finally, I define the predictive distribution as the distribution of a random variable $Y \equiv X + \mu$ where $X \sim F_X|_{x \in V(p^0, x^0|m)}$ for some assumed distribution of choices F_X , and $\mu \sim F_e|_{x,m}$ the effective distribution of the residuals in the new economic environment estimated from the residuals recovered in the projection. Formally, let $\{p^0, x^0\}$ be the new economic environment then,

Definition 7 ($F_X^{p^0,x^0}|m, x$) *Let F_X be an assumed distribution of choices provided by the econometrician.*[9] *Then, the distribution of choices over the Varian Supporting set is given by*

$$f_X^{p^0,x^0}|m(x) \equiv \frac{\mathbb{I}\left(x \in V\left(p^0, x^0|m\right)\right) f_X^{p^0,x^0}(x)}{\int_{V(p^0,x^0|m)} f_X^{p^0,x^0}}$$

where $supp\left(F_X^{p^0,x^0}\right) = B(p^0, x^0)$

Definition 8 ($F_{\widehat{e}}^{p^0,x^0}|m, x, v^0$) *Let $F_{\widehat{e}}$ be an estimation from recovered residuals 11 of the distribution given by equation 2, and let $v^0 \in V\left(p^0, x^0|m\right)$ be a demand vector in the supporting set. Then, the effective distribution of the error process in the new economic environment conditional on v^0 is given by $F_{\widehat{e}}^{p^0,x^0}|m, x, v^0$ consistent with equation 2, i.e.,*

$$f_{\widehat{e}}^{p^0,x^0}|m, x, v^0 = \begin{cases} f_{\widehat{e}|p^0\widehat{e}=0} & \textit{if } \widehat{e} \in int\left(A_{\widehat{e},v^0}\right) \\ \int_a dF_{\widehat{e}|p^0\widehat{e}=0} & \textit{if } \widehat{e} \in cl\left(A_{\widehat{e},v^0}\right) \textit{ and } a = argmin_{x \in A_{\widehat{e},v^0}} \|a - x\| \\ 0 & \textit{otherwise} \end{cases} \tag{5}$$

where $A_{\widehat{e},v^0} = \left\{e | v^0 + e \in \mathbb{R}_+^{L \times J} \text{ and } (p^0)'e = 0\right\}$

Definition 9 (Predictive distribution) *Given a data set $\left\{x^i, p^i\right\}_{i=1}^J$ the predictive distribution of choices y^0 for a new economic environment $\{x^0, p^0\}$, $F^0|_{m,x,p}(y^0)$, is defined as the distribution of a ran-*

[9]This distribution is provided by the researcher, by default it is set to be Uniform.

dom variable Y^0 where $Y^0 \equiv X^0 + \mu$ with $X^0 \sim F_X^{p^0,x^0}$ as in definition 7 and $\mu \sim F_e^{p^0,x^0}|m, x, X^0 = v^0$ as in definition 8.

Under assumptions 1 and 2 and consistency of the behavioral estimator, the predictive distribution in $B(x^0, p^0)$ is the expected distribution of choices. The predictive distribution conveys information over noise and fit of the model, choices and design. Intuitively, to accurately and precisely predict behavior in a new environment, the forecast should be made based on an estimated model that accurately describes the observed data (high goodness of fit) and that embeds information about underlying behavior.

Theorem 5 (Properties of the Predictive distribution) *Let assumption 1 and 2 hold. Then*

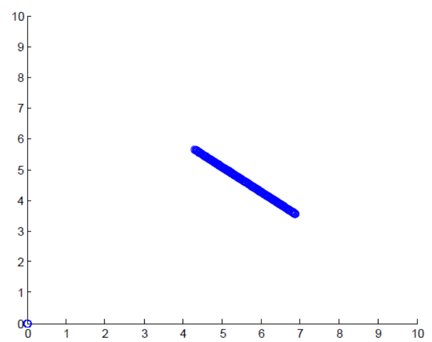
1. $\frac{\partial Var(Y^0)}{\partial \sigma_\varepsilon^2} \geq 0$
2. $\frac{\partial Var(Y^0)}{\partial J} \leq 0$
3. $\frac{\partial Var(Y^0)}{\partial q_i^0} \leq 0$ *where* $q^0 = \left[\int_{x \in B(p^0) \setminus B(p^i)} dF_X^{p^0,x^0}\right] \times \left[\int_{x \in B(p^i) \setminus B(p^0)} dF_X^{p^i,x^i}\right]$
4. *If observed data is consistent with GARP, then* $F^0|_{m,x,p} = F_X^{p^0,x^0}|m, x$
5. *Under the conditions in theorem 3, as* $F^0|_{m,x,p} \to F_e^{p^0,x^0}|m, x, v^0$ *where* m, v^0 *is the limiting demand.*

where $Y^0 \sim F^0|_{m,x,p}$

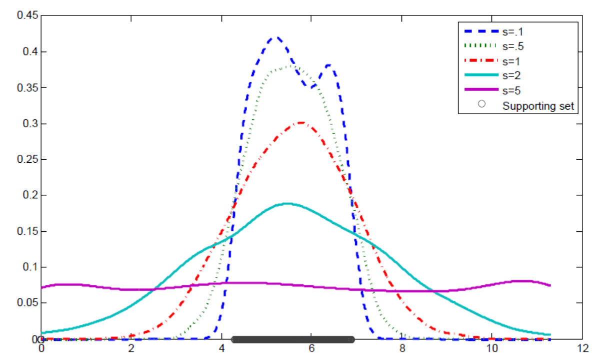
Proof. See Appendix D.1 ∎

The dispersion of the predictive distribution provides information on the noise of the underlying error process -1- and the amount of information that can be inferred from observed data to generate forecasts in the chosen economic environment. As more information is considered, these predictions become more precise, since tighter is the identification of the underlying rational component which implies that the supporting set shrinks -2-. If the relative distribution of budget sets is such that GARP imposes more demanding constraints involving $B(x^0, p^0)$, then the smaller the supporting set is expected and the less disperse the predictions are -3-. Finally, -5- provides an input to construct an asymptotically valid test statistic for the null of rationality from the residuals of the distribution. Notice that, even in the limit, the distribution any statistic based on the effective residuals does not have a standard distribution due to feasibility constraints.

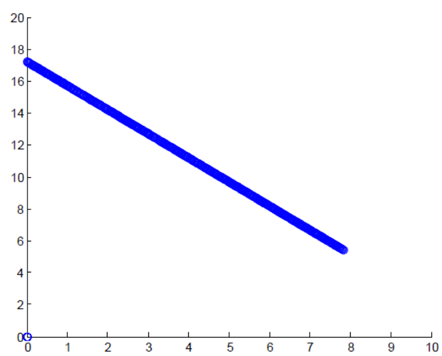
Figure 5 shows the effect of the size of the supporting set and σ_ε on the predictive distribution. Panel 5a shows the supporting set for a sequence of budget sets that impose demanding constraints on data from $B(p^0, x^0)$ to be rational, while panel 5c shows the case for a looser set of constraints. The relative size of the supporting set with respect to $B(p^0, x^0)$ is providing information about the falsifiability of the model given observed data, but does not provide information about the empirical



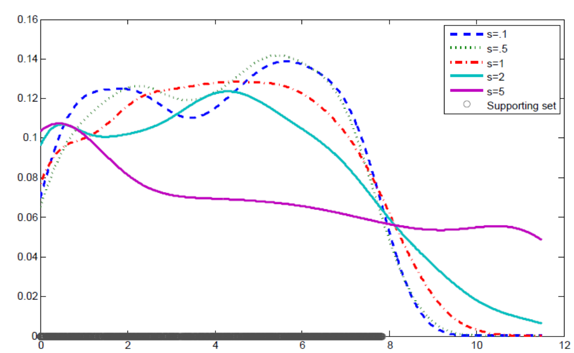
(a) Supporting set: demanding



(b) Predictive distribution for different σ given figure 5a



(c) Supporting set: not demanding



(d) Predictive distribution for different σ given figure 5c

Figure 5: The predictive distribution conveys information about fit and power

accuracy. When constructing the predictive distribution I incorporate the error process needed for observed data to consistent with the model. Panels 5b and 5d show the resultant predictive distribution once the error process is included, for different assumptions over the variance of the error process, where $\sigma_\varepsilon \in \{0.1, 0.5, 1, 2, 5\}$. It can be seen that, as the variance of the underlying error process increases the predictive distribution flattens out providing less information to predict choices in $\{x^0, p^0\}$.

I propose to measure the predictive accuracy by the information contained in the predictive distribution in the understanding that a better model delivers more precise/informative estimates. As measures of the tightness of the predictions I propose: (i) to measure the size of the smallest $\alpha-$ level credible set for predicted choices; and (ii) to measure the informational content of the predictive distribution as a measure of the information that can be inferred from data by imposing the model. The latter also allows for the construction of statistics to test whether the predictive distribution conveys significant more information than an uninformative prior.

3.5 Predictive ability measures

3.5.1 "Size" of the HPD $\alpha-$level credible set

A better model is one that delivers smaller credible sets for predictions, that are the result of more informative predictive distributions, that is, less disperse estimates. As shown is theorem 5, more precise forecasts are the result of higher fit and more demanding constraints imposed on data. Therefore I propose to gauge the performance of the model by the "size" of the smallest α-level credible set, the smaller the credible set the higher the predictive accuracy.

Define $C_\alpha^{p^0,x^0}$ as the $100(1-\alpha)\%$ highest posterior density (HPD) interval (set) given the predictive distribution defined by 16. I size this set by the ex-ante probability of feasible choices to be such that $x \in C_\alpha^{p^0,x^0}$, that is with respect to the distribution given by $F_X|\{x \in B(p^0, x^0)\}$. The proposed measure of the performance of the model is given by its complementary probability, that is ,the ex-ante probability that choices are not in the credible set, the higher this probability the smaller the credible set. The rationale for this "metric" for the size of the credible set is two fold: (i) it provides a well defined measure for the relative size of the confidence set since it is a well defined probability and (ii) if the researcher has prior information to assume that, before imposing the model, some alternatives are more likely to be chosen, this information should be accounted for when measuring informativeness of the predictions. Alternatively, one can measure the relative size of the complement of the credible set by the relative size of its euclidean norm with respect to the norm of the feasible set. If F_X is assumed to be uniform then the proposed measure can be understood as the relative size of the complement to the confidence set in the standard sense, i.e these two approaches are equivalent.

Definition 10 (Predictive Accuracy - Credible Interval) *Let $C_\alpha^{p^0,x^0}$ be the $100(1-\alpha)\%$ highest pos-*

terior density (HPD) credible interval (set)[10] *for the predictive distribution defined in 16, $F^0|_{m,x,p}$, then predictive accuracy measure based on the $\alpha-$level credible interval is defined as*

$$PA_{\alpha}^{p^0,x^0}|_{m,x,p} \equiv 1 - P_X\left(x \in C_{\alpha}^{p^0,x^0} | x \in B(p^0, x^0)\right) \tag{6}$$

Note that from proposition 4 the supporting set is convex, and from assumptions 1 and 2 the distribution of the error process in the interior of the positive quadrant is assumed to be continuous and unimodal, which implies that, unless $V\left(p^0, x^0|m\right) = B(p^0, x^0)$, the α-level confidence set is a meaningful construction to measure predictive accuracy.

Proposition 5 (Properties of Predictive accuracy index $PA_{\alpha}^{p^0,x^0}|_{m,x,p}$) *Assume 1 and 2 hold. Consider the predictive accuracy measure given by definition 10. Then,*

1. $PA_{\alpha}^{p^0,x^0}|_{m,x,p} \in [0,1]$
2. *If* $m = x$ *then* $PA_{\alpha}^{p^0,x^0}|_{m,x,p} = 1 - (1-\alpha)P_X\left(x \in V\left(p^0, x^0|m\right)|x \in B(p^0, x^0)\right)$.
3. $\frac{\partial PA_{\alpha}^{p^0,x^0}|_{m,x,p}}{\partial \sigma_{\varepsilon}} < 0$
4. $\frac{\partial PA_{\alpha}^{p^0,x^0}|_{m,x,p}}{\partial P_X(x \in V(p^0,x^0|m)|x \in B(p^0,x^0))} < 0$
5. *Given* σ_{ε}, *under the conditions in theorem 3, as data becomes dense* $(J \rightarrow \infty)$ $\lim_{J\rightarrow\infty} PA_{\alpha}^{p^0,x^0}|_{m,x,p} = PA^*$, *where* $PA^* \equiv \max PA_{\alpha}^{p^0,x^0}|_{m,x,p}$

Proof. See Appendix 3.5 ■

The proposed measure is interpreted as a measure of the size of the set of feasible choices that are not in the credible set. The more informative the predictive distribution, the smaller the credible set, the higher the proposed measure. If the measure is significantly higher than α then the model conveys more information than $F_X|x \in B(p^0, x^0)$. The latter is relevant when considering F_X as an uniform distribution since most of the goodness of fit and power literature relies, as an alternative hypothesis to rationality, on the definition of irrational behavior by Becker (1962) that translates into a uniform distribution over the feasible set. Proposition 5 shows that the proposed measure has the desirable properties, combining an assessment of fit and power as it is shown by (3) and (4). More precise

[10]The highest posterior density interval is an interval which has a probability of coverage $1-\alpha$ and has the minimal length. Formally,

Definition 11 (Highest posterior density(HPD) interval) *A $100(1-\alpha)\%$ HPD interval is a region that satisfies two conditions*

1. *The probability of content is* $100(1-\alpha)\%$
2. *The minimum density of any point within the region is equal to or larger than the density of any point outside the region*

estimates result in higher predictive accuracy -(3)- as well as more stringent economic environments result in higher measures through the effect on the size of the Varian supporting set -(4)-.

3.5.2 Information theory based measures

A model with high predictive accuracy is expected to deliver more informative predictive distributions. Then the performance of the model can be appraised by the gain in information to predict choices when considering the predictive distribution constructed from the model and observed data, with respect to the prior that does not incorporate any information from choices nor the structure prescribed by the model.

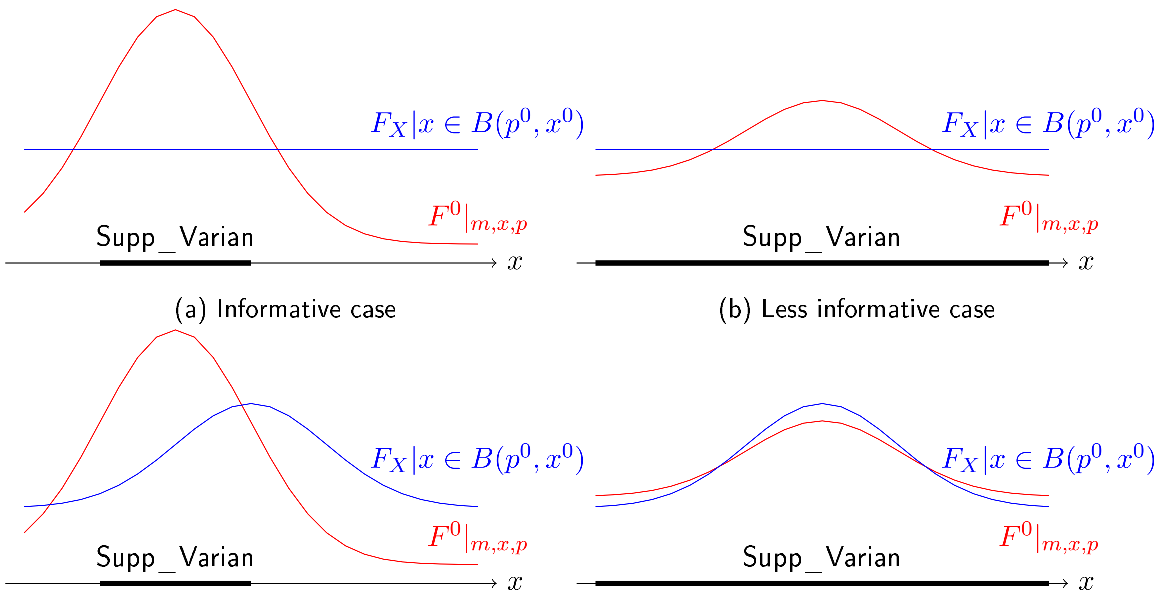
Let $F^1 \equiv F^0|_{m,x,p}$ be the predictive distribution constructed as in definition 16 and let $F^2 \equiv F_X|x \in B(p^0, x^0)$ be the prior distribution of choices over the new budget set. Consider the distance between these two distributions, if these two are close, then there is no gain in certainty by predicting using the model; on the other hand, if these two distributions are considerably different, then the model is providing significant information to construct predictions. More demanding constraints imposed on data result on predictive distributions that are more informative due to smaller supporting sets. Similarly, as higher the fit of the data to the model, the lower the variance of the estimated error process and the more informative the predictive distribution is and the higher the distance between the predictive and prior distribution of choices on $B(p^0, x^0)$.

Divergence measures provide a measure of the differential information embed in the predictive distribution with respect to the prior. Consider figure 6, panels 6a and 6b. The differential information measures are given by the distance between the two distribution functions in the figure, where the assumed prior is a uniform distribution over the feasible set. Then, as more informative the predictive distribution of choices is, bigger would be the difference between the two distributions and bigger the gain in information from the data by imposing the model. Obviously, this measure depends on the assumed prior distribution of choices, as it shown from the comparison between panels 6a-6b and 6c-6d.

The considered divergence measures are: (i) Relative entropy divergence measures; (ii) Hellinger distance; and (iii) total variation distance. These are defined in section 3.5. The connection between these measures have been established in the literature, see results in section 3.5.

Definition 12 (Predictive accuracy index - Information theory) *Let $F^1 \equiv F^0|_{m,x,p}$ be the predicted distribution of choices based on the model and observed data as given by definition 16 and $F^2 \equiv F_X|x \in B(p^0, x^0)$ be the prior (assumed) distribution of choices. Then,*

$$PA^{p^0,x^0}_{info-i}|_{m,x,p} \equiv D_i \left(F^1||F^2\right) \tag{7}$$



(a) Informative case

(b) Less informative case

(c) Informative case - Informative prior

(d) Less informative case - Informative prior

Figure 6: Predictive distribution of choices vs. prior distribution

where $i \in \{Kullback - Leibler, Hellinger, TV\}$[11]

It has been well established the connection between Kullback-Leibler and Shannon entropy, $PA^{p^0,x^0}_{info-KL}|_{m,x,p} = D_{KL}\left(F^1||F^2\right) = -H(f^1) - E_{f^1}\left(\ln f^2\right)$ where $H(\cdot)$ is the Shannon entropy. The entropy of a distribution can be understood as the uncertainty in terms of predicting an outcome from the distribution. Moreover, if $F_X|x \in B(p^0, x^0) \sim U_{B(p^0,x^0)}$, then $E_{f^1}\left(\ln f^2\right) = \ln\frac{1}{n} = -\ln n$ where $n = \int_{B(p^0,x^0)} dx$, therefore $PA^{p^0,x^0}_{info-KL}|_{m,x,p} = \ln n - H(f^1)$, that is, when the prior is uninformative maximizing $PA^{p^0,x^0}_{info-KL}|_{m,x,p}$ is equivalent to minimizing the entropy of the predictive distribution, i.e. the uncertainty to predict outcomes from the predictive distribution.

Proposition 6 (Properties of Predictive accuracy - Information theory) *Let assumptions 1 and 2. Consider the Predictive Ability measures as given by definition 12. Then,*

1. $PA^{p^0,x^0}_{info-i}|_{m,x,p} \in [0,1]$ *for* $i \in \{Hellinger, TV\}$, $PA^{p^0,x^0}_{info-KL}|_{m,x,p} \geq 0$. *Moreover,* $PA^{p^0,x^0}_{info-i}|_{m,x,p} = 0$ *if and only if* $F^1 = F^2$ *for* $i \in \{KL, Hellinger, TV\}$

2. *If* $\widehat{x} = x$ *(*$\widehat{\varepsilon}_j = 0$ *for all* j*), then* $Dom(F^1) = V\left(p^0, x^0\right)|m$. *Furthermore, if the prior* F_X *is uniform then*

   (a) $PA^{p^0,x^0}_{info-KL}|_{m,x,p} = -\ln\gamma$

   (b) $PA^{p^0,x^0}_{info-Hell}|_{m,x,p} = \sqrt{2}\left(1 - \sqrt{\gamma}\right)$

   (c) $PA^{p^0,x^0}_{info-TV}|_{m,x,p} = 1 - \gamma$

[11] The formal definitions are provided in Appendix 3.5

where $\gamma = \int_{V(p^0,x^0|m)} f^2(x)dx$

3. $\frac{\partial PA^{p^0,x^0}_{info-i}|_{m,x,p}}{\partial \sigma_{\widehat{\varepsilon}}} < 0$, *for* $i \in \{KL, Hellinger, TV\}$

4. $\frac{\partial PA^{p^0,x^0}_{info-i}|_{m,x,p}}{\partial P_X(x \in V(p^0,x^0|m)|x \in B(p^0,x^0))} < 0$ *for* $i \in \{KL, Hellinger, TV\}$

Proof. See Appendix D ■

Proposition 6 summarizes some of the desirable properties of a predictive accuracy index. These indices are zero if the data conveys no distinct information than the prior distribution of choices, and increases as the information extracted from data by imposing the model results on a predictive distribution that is more and more different from the prior distribution of choices -(1)-. If the observed data is perfectly rational, the predictive distribution is given by the truncation of the prior distribution to the Varian supporting set, reflecting the relative size of this set with respect to the feasible set, γ as defined above, -(2)-. Moreover, these measures worsen as worse is the fit of the model -(3)-, or weaker are the constraints imposed by the model-(4)-.

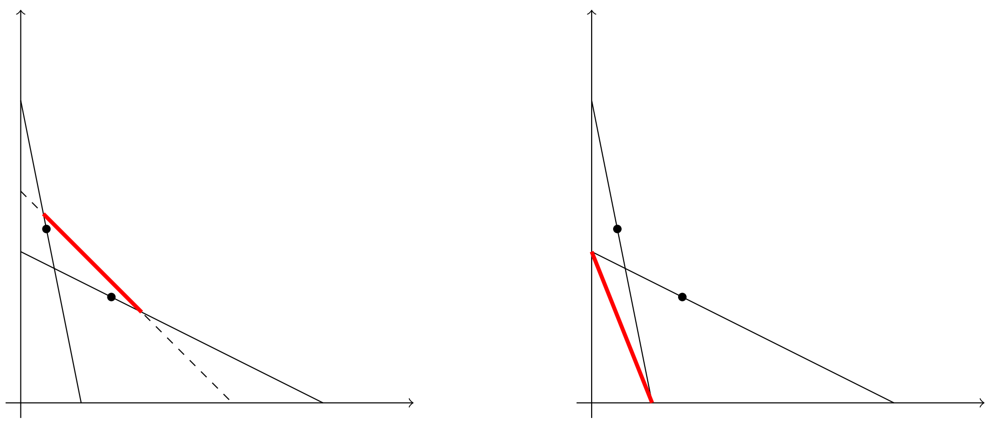
Constructing the predictive distribution not only allows the researcher to construct measures of the quality of the model as in the previous section, but also to test whether this distribution conveys statistically different information than an uninformative prior. Consider the following hypothesis testing problem

$$H_0 : F^1 = F^2 \text{ vs. } H1 : F^1 \neq F^2 \tag{8}$$

If H_0 is not rejected, then the prior and predictive distribution do not convey significantly different information, thus the model does not provide any additional information about behavior. The above hypothesis can be tested by implementing a two-sample Kolmogorov-Smirnov test. Different test statistics have been proposed in the literature to test, nonparametrically, for the equality between two distributions; refer to Pardo (2005) for the details.

3.5.3 Relevance of $B(p^0, x^0)$

Notice that the definition of the set of predicted choices depends crucially on the economic environment in which the prediction is performed, as shown in theorem 5. This is not relevant if the prediction is realized over budget sets that are relevant for the researcher; but if the aim is to assess the performance of the utility maximization model it needs to be noticed that the size of the supporting set is negatively and strongly correlated to the probability of finding GARP violations on the new budget set. This is particularly important when the comparison is made across individuals that faced different $\mathbf{B}(p, x)$. For example consider the case of figure 7, where only $B(p^0, x^0)$ has changed from panel 7a to 7b. A more suitable assessment for the quality of the model is the expected value of the proposed measures across different realizations of $\{p^0, x^0\}$. Unless the comparison is made across a sufficiently dense sequence of budget sets, this dependency will encumber the interpretation of the results when



(a) $B(p^0, x^0)$ demanding for prediction

(b) $B(p^0, x^0)$ not demanding

Figure 7: Predicting consumer's behavior - $B(p^0, x^0)$ matters

comparing individuals that faced different economic environments.

Alternatively, consider a predictive accuracy index where the forecast is performed on the observed economic environments. Consider a leave-one-out prediction, that is, for each j, consider the set of observations $J \setminus j$ to construct the predictive distribution for $B(p^j, x^j)$; and then these measures are summarized across observations. This type of measure would have similar properties to the ones established above, but overcoming the dependency with respect to $B(p^0, x^0)$ by relying exclusively on the data; reinforcing the concept that the quality of the model is a feature not only of the model itself but of the data which behavior aspires to replicate.

3.6 Finite sample properties

In finite samples, GARP identifies the set of demand vectors that are rational conditional on observed economic environments. As new observations are revealed from budget sets that impose demanding constraints on data to be rational, the set of consistent demands shrinks, as shown in Theorem 4. Figure 14 in Appendix 5.2 shows how for a perfectly well-behaved decision maker, the set of choices from $B(p^0, x^0)$ that are consistent with $m \in \mathbf{M}(\mathbf{B}(p, x))$ shrinks as the number of observations increases, for all $m \in \mathbf{M}(\mathbf{B}(p, x))$. This phenomenon does not only depend on the number of observations but also on the particular distribution of $\{B(x^i, p^i)\}_{i=1}^{J}$ and sequence of choices.

Assumptions 1 and 2 imply that the error process is partially identify since

$$\begin{aligned} \widehat{\varepsilon} &= x - \arg \min_{\widetilde{m} \in \mathbf{M}(\mathbf{B}(p,x))} \|x - \widetilde{m}\| \\ &\leq x - m \equiv \varepsilon \end{aligned}$$

where the last inequality is strict in the case where $\arg\min_{\widetilde{m} \in \mathbf{M}(\mathbf{B}(p,x))} \|x - \widetilde{m}\| \neq m$. In particular, as

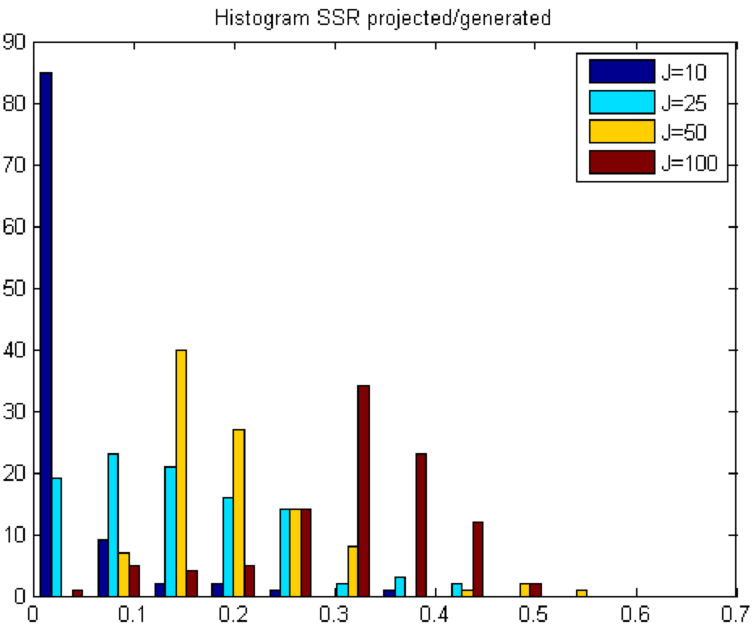


Figure 8: Histogram of $\frac{Projected-SSR}{Generated-SSR}$ for $J = 10, 25, 50, 100$, and 100 repetitions. Choices where generated from a Cobb-Douglas utility function $u(x, y) = \sqrt{xy}$ plus ε given assumptions 1 and 2, where $\varepsilon \sim N(0, 5)$. $\mathbf{B}(p, x)$ was generated such that $p_1 \sim U[5, 10]$, $\frac{p_2}{p_1} \sim U[.5, 2]$ and $w \sim U[80, 120]$.

the number of observations increases, $\mathbf{M}(\mathbf{B}(p, x))$ shrinks, and therefore $\widehat{\varepsilon} \to \varepsilon$.

The relevance of this effect is that the underestimation of the error for any finite sample size prevents the econometrician from recovering a suitable distribution for the error process to construct the predictive distribution. The identified error is still the object of interesting when measuring the distance of the data to the model. Consider figure 8, where $\varepsilon \sim N(0, 5)$ and $J = 10, 25, 50, 100$. The underestimation of the residual process is significant for all considered sample sizes and he gap between $\|\varepsilon\|$ and $\|\widehat{\varepsilon}\|$ is jointly defined by the error process and the sample size, given a DGP for $\{(p^i, x^i)\}_{i=1}^{J}$. These features are consistent across DGP, see figures 15 and 16 in Appendix 5.2.

Given some parametric assumption on the underlying distribution of the error process, F_{ε}^{θ}, I propose to recover the sufficient statistics of such distribution (for θ) by simulating the error process given assumptions 1, 2 and F_{ε}^{θ} for different values of θ in the observed economic environments, recovering $\widehat{\varepsilon}(\theta)$ and comparing its distribution to the distribution of $\widehat{\varepsilon}$ recovered from data. I choose θ to be the one that generates a distribution that is the closest to the empirical distribution of $\widehat{\varepsilon}$, i.e., $\widehat{\theta} = \arg\min d\left(F_{\varepsilon}^{\theta}, F_{\varepsilon}\right)$.

Estimating $\widehat{\theta}$ as above implies computing the projection for each of the repetitions in the simulation which is computationally really expensive. Alternatively, I propose a series of algorithms to generate F_{ε}^{θ} without computing the projection in each iteration but relying on the structure of the problem and the

σ	Algorithms - σ^2_ε		
	1	2	3
0.5	0.8662	0.9098	0.5348
1	0.5910	0.6544	**0.1320**
2	**0.5034**	**0.5294**	0.5574
5	0.9312	0.9406	0.8928
10	0.9448	0.9448	0.9190

Table 1: Kolmogorov-Smirnov two tailed test statistics with respect to measure constructed based on the projection for $\sigma = 3$ for two alternative distributions over a CD utility function.

performed projection to simulate the set $\mathbf{M}(\mathbf{B}(x, p))$.

Algorithm 1 is based on the assumption that $\widehat{x} = m$. Given the discussion in the previous section, for small samples one may argue this is a particularly strong assumption. Alternative, algorithm 2 constructs the consistent set $\mathbf{M}(\mathbf{B}(p, x))$ by intersecting, in a sequential manner, the Supporting set; that is pick i at random and construct $V(p^i, x^i|\widehat{x}_{-i})$, then $k \neq i$ and construct $V(p^k, x^k|\widehat{x}_{-\{i,k\}})$ and do so until the set $\mathbf{J} = \{1, \ldots, J\}$ is exhausted. Algorithm 3 does not rely on the projected demand vector but simulates $\mathbf{M}(\mathbf{B}(p, x))$ sequentially. The details for the construction of these algorithms are presented in Appendix E. In table 1 the results are presented under the assumption that $F_\varepsilon = \sigma\Phi(\varepsilon)$, where Φ is a standard normal distribution. The results show that the proposed algorithms produce good estimates for σ^2, algorithm 1 and 3 underestimate the variance while algorithms 2 produce unbiased estimates.

4 Comparison to other measures in the literature

4.1 Comparison to goodness of fit measures

The proposed measures reflect the dispersion of the predictive distribution that depends on the fit of the model (distribution of $\widehat{\varepsilon}$) and the number and relative distribution of budget sets through their effect on the supporting set; as shown in theorem 5. Naturally, these measures are positively correlated to measures of fit previously proposed in the literature, but they do not produce the same ordinal result in the comparison across data sets. In particular,

$$\frac{\partial PA_i|\mathbf{B}(x,p)}{\partial\|\widehat{\varepsilon}\|} < 0$$

but

$$\|\widehat{\varepsilon_1}\| > \|\widehat{\varepsilon_2}\| \nRightarrow PA^1_i|\widehat{\varepsilon_1} < PA^2_i|\widehat{\varepsilon_2}$$

where $\|\widehat{\varepsilon}\|$ measures the distance from the data to the model. The last inequality does not hold since, even if noisier, the economic environments faced by individual 1 may have exposed her to more demanding constraints what, even if behavior seems to be further from the set of consistent choices, more information is extracted from observed behavior which further constraints the supporting set.

To show that this is the case, consider again figure 4, panels 4a and 4b. In this case $\widehat{\varepsilon} = 0$ for both panels, but panel 4a is more informative to predict behavior than panel 4b; similar in the case of panels 4c-4d. That is, the size of the supporting set plays a significant role in the predictive distribution, independently of the recovered residuals. The result is stronger, it can be the case that, models with worse fit deliver stronger predictions. Consider now figure 5 and compare the predictive distribution in panels 5b and 5d. Even when $\sigma_\varepsilon = 2$ in panel 5b, this predictive distribution is more informative than the predictive distributions in panel 5d for all $\sigma_\varepsilon < 2$. The proposed measures of predictive ability are correlated to the measures proposed in the literature but do not provide the same information and need not generate the same ordinal results when comparing across data sets.

4.2 Comparison to power approaches

The standard approach in the literature is to compare the results for the goodness of fit measures to the results that would be obtained if data were generated randomly from a uniform distribution under the assumption of Becker (1962) irrational behavior[12]. Beatty and Crawford (2011) proposes to adjust a measure of success, or fit, with a measure of the target area to account for power. The target area is defined as the probability of choices being rational if they were to be generated by a random subject, that is, it is an ex-ante measure of stringentness of the model. The predictive ability measures proposed in this paper account for power though the effect of the relative distribution of $\mathbf{B}(p, x)$ on the size of the supporting set. As discussed in section 6.2, the size of the supporting set does not only depend on the relative distribution of budget sets, but also in the interaction between such distribution and observed choices; that is the stringentness that observed budget sets imposed in observed choices, not in an average expected choice if choices were to be generated at random. Consider again figure 4, Beatty and Crawford (2011) type of measure account for the effect of $\mathbf{B}(p, x)$ on power for the case of the comparison between 4c and 4d but would not discern between the cases in figure 4a and 4b.

The proposed measures account for fit and power but do not deliver the same qualitative results as other measures that proposed to do so, since the tradeoff proposed here is conditional on observed choices. Andreoni and Harbaugh (2013) proposes a conditional approach for power measures but do not provide a meaningful way to combined with fit, that is, their interpretation should be made conditional on fit.

[12]Andreoni and Harbaugh (2013) proposes a series of power indices that account for information about observed choices to define power. They do not deal on the particular manner to combine fit and power.

5 Empirical Performance

The results from section 3 state that is expected that as the number of observations increases the size of the supporting set reduces (proposition 4) as well as the variance of the predictive distribution (theorem 5) and consequently the predictive accuracy measures improve. Also, under certain conditions on the distribution of the budget sets, it is expected that the predictive distribution converges to the distribution of the effective residual process. For a given sample size and $\mathbf{B}(p,x)$ it is expected that the variance of the predictive distribution increases with σ_{ε}^2, and therefore the predictive ability worsen. Finally, if the relative distribution of budget sets imposes less stringent constraints in data, the predictive ability measure are expected to be worse.

5.1 Experimental Data

5.1.1 Data

In this section I use data from Choi *et al.* (2007a). This data set was obtained from a series of experiments designed to study decisions under uncertainty. The 93 subjects in the experiment are presented with a graphical interface displaying standard, two dimensional, budget constraints on the screen. The experiment was conducted at the Experimental Social Science Laboratory at UC Berkeley and each session consisted of 50 independent decision rounds.

Table 6 shows the summary statistics for the proposed measures and other standard measures of fit that are standard in the literature.[13] The predictive accuracy measures are computed for the observed economic environments in the regime of leave-one-out prediction, as explained in section 3.5.3. Table 7 shows the correlations among these measures, standard measures in the literature, other alternative measures of fit constructed from the residuals of the projection[14] and a measure of the ex-ante mean probability of observing violations to the model given observed economic environments as a proxy for the stringentness of the constraints imposed by the model. Almost 20% of the subjects are rational, and standard measures of fit (raw R^2, Afriat and Varian measures) are high, consistent with the fact that, given observed budget sets, the ex-ante probability of detecting violations to the model is low. Moreover, the correlation between these measures and an estimation of this ex-ante probability is negative.

5.1.2 Empirical performance PA measures and comparison with other measures

Tables 8-12 show the impact of measures of fit and ex-ante power on the proposed measures. Table 8 presents the results for the predictive accuracy measure given by definition 10. The results are as expected given proposition 5, i.e. the proposed measure responds positively to changes in the raw fit of the model (a measure of $-\|\widehat{\varepsilon}\|$) and therefore negatively to changes in σ_{ε}. Moreover, when

[13] The consider measures of fit are the proposed by Afriat (1972a), Varian (1990), Houtman and Maks (1985) as well as the number of violations to WARP and GARP.

[14] The considered measures are raw R^2, adjusted R^2 and weighted R^2. Refer to appendix F.1 for the formal definition of these measures.

considering designs with an ex-ante probability of detecting violations above 2.5% this measure also improves as the design becomes more stringent. Table 9 shows that, when controlling for the R^2 of the projection exercise, standard measures proposed in the literature do not have a significant effect on the predictive accuracy of the model.

Tables 10-12 present the analysis for the measures given by definition 12. The predictive accuracy measure based on the Kullback-Leibler discrepancy measure shows a quadratic dependency with respect to the measure of raw fit (R^2) while the ex-ante probability of detecting violations has a positive but insignificant effect on the predictive ability. The Hellinger and the Total Variation measures show similar dynamics, consistent with the results from proposition 6, tables 11 and 12 respectively in Appendix F. Notice that, the Hellinger measure is more sensitive to fit, while the total variation measure shows more sensitivity to changes in power, measured as the ex-ante probability of detecting violations to the model given the observed economic environments. In all cases, once I control for fit, standard measures proposed by the literature show negative[15] effect on the ability to produce accurate predictions, which is consistent with the negative correlation between these measures and how demanding are the observed designs as it is shown in table 7.

As discussed in section 4.1, the predictive accuracy measures not only account for fit but also for the effect of the distribution of $\{p^j, x^j\}_{j=1}^{J}$ on the ex-ante power of the revealed preference test. Consider subjects 209 and 606 whose choices are displayed in figure 11 in Appendix F. Common measures used in the literature imply that subject 209, in panel 11a seems to be closer to the rationality benchmark. A closer analysis of the data suggests that subject 606 has been exposed to more demanding constraints given the budgets sets that were presented to her, where the mean probability of finding a violation is 13.49% by comparison to 1.79% for subject 209. The forecasting ability measures proposed in this paper account for this effect showing that, in effect, despite of more noise, more information can be inferred from choices for subject 606 than for 209, as shown in table 2.

Other measures that combine fit and power have been proposed in the literature, refer to section 8. The proposed methodology provides a meaningful trade off for these two components and also incorporates the information contained in observed choices, as discussed in section 4.2. It is standard use in the literature to adjust by power computed as the performance of choices if they were to be generated by uniformly random behavior, but Andreoni and Harbaugh (2013) and Dean and Martin (2013) propose alternatives that exploit further information from observed data to compute the power of the test. Consider subjects 209 and 603 in table 3, even when standard goodness of fit measures and adjusted ones (Adjusted R^2 BC) would indicate that the performance of the model is better for subject 209, subject 603 generates a more informative predictive distribution.

[15] Only significant for the KL measure.

	Subject 209	Subject 606
R^2	0.9740	0.9306
Adj R^2	0.0009	0.1878
FA CI (95%)	0.4557	0.5175
FA KL	0.3537	0.5280
FA Hellinger	0.2076	0.2910
FA TV	0.3372	0.4173
WARP	15	18
GARP	94	241
Afriat	0.9290	0.8390
Varian	0.8250	0.4700
HM	46	44
$\overline{P(viol)}$	0.0179	0.1349

Table 2: Data for subjects 209 and 606, Choi *et al.* (2007a)

	Subject 603	Subject 209
R^2	0.8597	0.9740
Adjusted R^2 BC	-0.1217	-0.0251
Afriat	0.6860	0.9290
FA CI (95%)	0.5782	0.4556
FA KL	1.1049	0.3537
FA Hellinger	0.5841	0.2076
FA TV	0.5916	0.3372

Table 3: Data for subjects 209 and 603, Choi *et al.* (2007a)

	Mean - Full	Mean - Half	|Test-statistic|	df $\approx$	P-value
R^2	0.9906	0.9989	2.0747**	154	0.0397
Adj R^2	0.1072	0.1100	0.2417	155	0.8093
Weighted R^2	0.9906	0.9969	2.1119**	156	0.0363
Mean ex-ante prob	0.1142	0.0788	3.6568***	153	0.0004
FA CI (95%)	0.4935	0.3989	9.1299***	143	0.0000
FA KL	0.3072	0.1915	6.3379***	91	0.0000
FA Hellinger	0.2962	0.2162	5.2035***	149	0.0000
FA TV	0.3940	0.3040	8.0911***	147	0.0000

Table 4: Data for 81 subjects, Choi *et al.* (2007a), test for different in means Welch's t-test.

5.1.3 Effect of the number of observations

To analyze the effect of the number of observations in the amount of information that can be extracted from data to generate predictions I compare the results when considering the first 25 observations and the full sample. Assuming that individual preferences are stable along the experiment, the comparison is made between data set that were generated by the same DGP. The results are presented in table 4. The number of rational individuals jumps from almost 20% to 50%. When comparing the results for half and full sample, the fit of the model is significantly higher when considering 25 observations, but these results do not imply changes in behavior but the fact that, with fewer observations, observed behavior is more likely to satisfy the constraints imposed by the model, regardless of the true underlying behavior. In particular the ex-ante mean probability of detecting violations is 7.88% when considering 25 observations, while the model is significantly more stringent when considering the full sample, 11.42%. Consistent with the results from propositions 5 and 6, the ability to produce accurate predictions increases. This is due to the amount of extra information that can be inferred from underlying behavior when more choices are observed, even when more noise increases (R^2 of .9906 with respect to .9989).

Table 13 in Appendix F displays the effect of changes in fit and power on changes in the performance of the model measured by its predictive accuracy, when the sample size is increased from 25 to 50 observations. For the measure based on the relative size of the confidence interval, columns (1)-(2), changes in fit and power are significant to explain the observed changes in predictive accuracy. For the measures based on the relative information contained in the predictive distribution, changes in fit have a negative or insignificant effect on the performance change.

5.1.4 Comparison to Parametric Recoverability

The proposed methodology allows the researcher to impose the minimal constraints needed to ensure the existence of a well behaved utility function, avoiding any type misspecification error induced by additional assumptions. If we were to assume a parametric specification for demand, the recovered residuals can be decomposed into a inconsistency and a misspecification component. Table 5 shows the effect of various parametric assumptions on the sum of square residuals for the whole sample and divided into rational and irrational subjects, where the considered parametric assumptions are: Cobb-Douglas,

	GARP	$\min\{CD, Leont, Lin\}$	$\frac{GARP}{\min\{CD,Leont,Lin\}}$
All subjects	357.54	18226.46	0.0289
	(1119.23)	(20018.90)	(0.1069)
Rational	0.00	17623.28	0.0370
	(0.00)	(25569.78)	(0.1889)
No Rational	491.62	18452.65	0.0258
	(1287.05)	(17483.49)	(0.0479)

Table 5: Effect of various parametric specifications on the mean of the sum of squared residuals, standard deviation in (). The estimation for the parametric specifications was performed by nonlinear least squares on the individual basis, by assuming behavior consistent with Cobb-Douglas, Leontief and Linear utility functions, considering the best fit per subject.

Leontief and linear[16], allowing for heterogeneous behavior across individuals.

5.2 Monte Carlo Results

Unless otherwise stated, the economic environments were generated such that $p_x \sim U[5, 10]$, $p_y = p_x \times a$ where $a \sim U[.5, 2]$ and $w \sim U[80, 120]$; this implies a mean probability of detecting a WARP violation of 7%, and that the probability of detecting at least one violation converges rapidly to 1.[17].

5.2.1 Convergence results

The consider DGP is consistent with assumptions 1 and 2, where the rational component was generated as the result of the maximization of $u(x, y) = \sqrt{xy}$ and $F_\varepsilon \sim N(0, 1)$. A total of 200 samples were created of (nested) 10,25,50 and 100 observations each.

Consistent with the discussion in section 3.6, the recovered residuals are expected to converge to the effective underlying error process as in equation 2 as the number of observations increases. Figure 9 shows that such convergence occurs in the data and it is slow, even when considering 100 observations it is far from full convergence.

When the number of observation increases, more information can be extracted from data resulting in higher predictive ability measures. Table 14 and figure 18 in Appendix F show that the proposed measures

[16]The correspondent assumptions are

$$CD \;\; u(x,y) = x^{\alpha}y^{1-\alpha} \;\; therefore \;\; x^* = \frac{\alpha w}{p_x} \;\; and \;\; y^* = \frac{(1-\alpha)w}{p_y}$$

$$Leontief \;\; u(x,y) = \min\{\beta x, y\} \;\; therefore \;\; x^* = \frac{w}{p_x + \beta p_y} \;\; and \;\; y^* = \frac{\beta w}{p_x + \beta p_y}$$

$$Linear \;\; u(x,y) = \gamma x + y \;\; therefore \;\; x^* = \frac{w}{p_x} \;\; and \;\; y^* = 0 \;\; if \;\; \gamma \frac{px}{py} \;\; and \;\; x^* = 0 \;\; and \;\; y^* = \frac{w}{p_y} \;\; if \;\; \gamma < \frac{p_x}{p_y}$$

[17]$\mathbb{P}$[at least one violation to WARP given n obs] $= 1 - (1 - \mathbb{P}_{avg}$[violation to WARP]$)^{\binom{n}{2}} \approx 1$ when $n \geq 12$

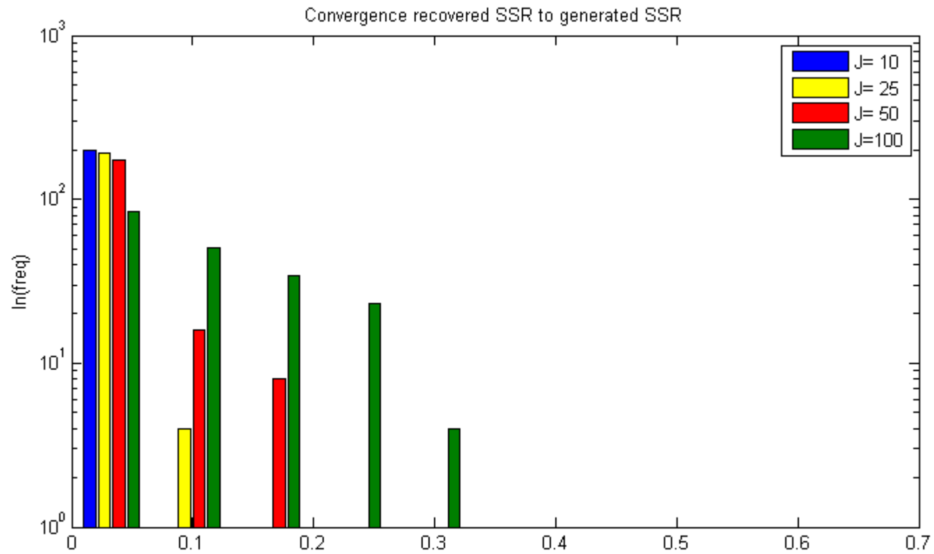


Figure 9: Histogram of $\frac{SSR_{recovered}}{SSR_{generated}}$

are consistent with the results from propositions 5 and 6 as the number of observations increases.

5.2.2 The effect of changes in the DGP

The DGP considered in this section is similar the one in section 5.2.1, but $J = 40$ and the possible values of sigma are given by $\sigma \in \Sigma = \{.5, 1, 2, 5\}$.

As expected, when the ratio of signal to noise increases, controlling for the economic environment, fewer subjects result to be rational, and we observed an significant increase in the computational cost. Table 15 in Appendix F shows the result for all the convergent simulations for each particular assumption over σ. The simulations results confirm the theoretical results stated in propositions 5 and 6, i.e. the measures deteriorate as σ increases, as well as the results become more disperse. Figure 19 in Appendix F shows the effect of changes in σ over the predictive ability defined in 10.

6 General Models of Economic Behavior

6.1 Set up

The proposed framework can be easily extended to a general class of economic models. Consider the case where the objective is to assess the quality of a given model M. Assume that the econometrician observes the behavior of an agent in J different economic environments. The features of the particular economic model(s) considered defined the characteristics of the economic environments, feasible behavior and consistent behavior with the model. Let X be the space of alternatives, then the economic environment realized defines the set of feasible alternatives for a decision maker in this context. Let

$Z(\theta^j)$ be the set of feasible alternatives in economic environment $\theta^j \in \Theta$ $j = 1, \ldots, J$, and let $Z : \Theta \to X$ be the feasible correspondence. I adopt the following notation: $y \equiv \left(y^1, \ldots, y^J\right)$ and $\mathbf{Y}(\cdot) \equiv \prod_{j=1}^{J} Y(\cdot)$, it follows that $\mathbf{Z}(\theta) = \prod_{j=1}^{J} Z(\theta^j)$ is the set of feasible alternatives vectors.

The model is assumed to be given by a correspondence $\mathbf{M} : \mathbf{Z}(\theta) \to \mathbf{Z}(\theta)$ that maps from the set of feasible choices to the subset of feasible choices that is consistent with the axiomatic model, that is

$$\mathbf{M}(\mathbf{Z}(\theta)) \equiv \{m \in \mathbf{Z}(\theta) : m \text{ satisfies model M}\}$$

Observed choices, $x(\theta)$, are assumed to be generated by the model $\mathbf{M}$ and an additive idiosyncratic error process ε; that is $x(\theta) = m(\theta) + \varepsilon$, where $m(\theta) \in \mathbf{M}\left(\mathbf{Z}(\theta)\right)$ and $x(\theta) \in \mathbf{Z}(\theta)$, i.e. observed choices must be feasible. Imposing feasibility may constrain the domain of the effective error process to be such that choices generated by the model and observed choices lie on $\mathbf{Z}(\theta)$. Without imposing any assumption on the nature of the feasible correspondence, I assume that latent choices are generated by the model and a well behaved error process, but observed choices are given by the closest feasible alternative. Formally, let e be the effective error process then,

$$dom\left(e|m(\theta)\right) \equiv \{e : (m(\theta) + e) \in \mathbf{Z}(\theta) \text{ and } m(\theta) \in \mathbf{M}\left(\mathbf{Z}(\theta)\right)\} \equiv \prod_{j=1}^{J} \left(Z\left(\theta^j\right) - m_j\left(\theta\right)\right)$$

Consider the following assumptions,

Assumption 4 (DGP) *Data is generated by a latent process given by*

$$x^*(\theta) = m(\theta) + \varepsilon$$

where $m(\theta) \in \mathbf{M}\left(\mathbf{Z}(\theta)\right)$. Then, observed data is given by

$$x(\theta) = \begin{cases} x^*(\theta) & \text{if } x^*(\theta) \in \mathbf{Z}(\theta) \\ argmin_{a \in Z(\theta)} \|x^*(\theta) - a\| & \text{otherwise} \end{cases} \tag{9}$$

where $x(p) \in \mathbf{B}(p, x)$ and $e \sim F_{\varepsilon|m(\theta),\mathbf{Z}(\theta)}$ is the effective error process given equation (9), i.e

$$e \text{ is such that } x(\theta) = \tilde{m}(\theta) + e \text{ for some } \tilde{m}(\theta) \in \mathbf{M}\left(\mathbf{Z}(\theta)\right)$$

Assumption 5 (Unconstrained Error Process) *The unconstrained process ε is assumed to be i.i.d. with continuous and symmetric p.d.f. such that $E(\varepsilon) = F_{\varepsilon}^{-1}\left(\frac{1}{2}\right) = 0$, $\frac{\partial f(\varepsilon)}{\partial \varepsilon}\varepsilon_j < 0$ for all $\varepsilon_j \neq 0$ and $\varepsilon \amalg m$, for some $m \in \mathbf{M}\left(\mathbf{Z}(\theta)\right)$.*

Assumptions 4 and 5 jointly define the distribution of the observed error process. These assumptions are agnostic in terms of the particular process that generated the data, since it relies uniquely on the conditions imposed by the axiomatization provided by the theory, preventing overrejection due to misspecification error. Moreover, the error process is assumed to be additive and, prior to the imposition of feasibility constraints, independent of the model, avoiding the imposition of any further structure to the problem.

6.2 Predictive Distribution

The predictive distribution is constructed as the expected distribution of choices on a new economic environment, given assumptions 4, 5 and observed data. I start by assuming the availability of an estimator of the behavior function, i.e. $\widehat{m}(\theta)$ given observed θ and $x(\theta)$.

Given a vector of consistent behavior, the axiomatic model prescribes the set of alternatives that are consistent with the model in the new economic environment, the supporting set, which is inherit from GARP. The estimation exercise delivers a behavior estimate that is consistent with the model based on observed data, that can be used to construct the supporting set in a new economic environment.

Definition 13 (Supporting set) *Let $m \equiv \left(m^1, \ldots, m^J\right)$ be a vector of choices consistent with the model, i.e. $m \in \mathbf{M}(\mathbf{Z}(\theta))$. Given a new economic environment $Z\left(\theta^0\right)$, the supporting set is defined as*

$$S\left(\theta^0|m\right) = \left\{x^0 \in Z(\theta^0) | \left\{m, x^0\right\} \in \mathbf{M}\left(\mathbf{Z}\left(\{\theta, \theta^0\}\right)\right)\right\} \tag{10}$$

Generally, new economic environments reduces the cardinality of the set of consistent behavioral functions, the relative size of the predicted area depends on the number of observations. If this is the case the relative distribution of observed and newly considered economic environments and the estimated behavioral functions. In particular, the supporting set is expected to be smaller as bigger (and more demanding) the data set is.

The distribution of this error process can be estimated from the residuals from the estimation of the consistent behavioral function, provided that feasibility conditions are incorporated. The error is defined as the residual from the estimation procedure, i.e.

$$\widehat{e} = x(\theta) - \widehat{x} \tag{11}$$

where $\widehat{x}$ is an estimator for $m(\theta)$. Given a parametric assumption on F_ε, the relevant parameters can be estimated by maximum likelihood by imposing the structure induced by assumptions 4 and 5; for small samples one can modify accordingly the algorithms proposed in section 3.6

Finally, I define the predictive distribution as the distribution of a random variable $Y \equiv X + \mu$ where $X \sim F_X|_{x \in S(\theta^0|m)}$ for some assumed distribution of choices F_X , and $\mu \sim F_e|_{x,m}$ the effective distribution of the residuals in the new economic environment estimated from the residuals in the estimation of the behavioral function. Formally, let $\{\theta^0\}$ be the new economic environment then,

Definition 14 ($F_X^{\theta^0}|m,x$) *Let F_X be an assumed distribution of choices provided by the econometrician.*[18] *Then, the distribution of choices over the Supporting set is given by*

$$F_X^{\theta^0}|m \equiv F_X\left(x|x \in Z(\theta^0) \cap S\left(\theta^0|m\right)\right)$$

Definition 15 ($F_e^{\theta^0}|m,x,s^0$) *Let $F_{\widehat{e}}$ be an estimation from recovered residuals 11 of the distribution given assumptions 4 and 5, and let $s^0 \in S\left(\theta^0|m\right)$ be a vector of behavior in the supporting set. Then, the effective distribution of the error process in the new economic environment conditional on s^0 is given by $F_{\widehat{e}}^{\theta^0}|m,x,s^0$ consistent with assumptions 4 and 5.*

Definition 16 (Predictive distribution) *Given a data set $\left\{x^i,\theta^i\right\}_{i=1}^J$ the predictive distribution of choices y^0 for a new economic environment θ^0 is defined as $F^0|_{m,x,\theta}(y^0) = P(Y^0 \leq y^0)$ where $Y^0 \equiv X^0 + \mu$ with $X^0 \sim F_X^{\theta^0}$ as in definition 14 and $\mu \sim F_e^{\theta^0}|m,x,s^0 = X^0$ as in definition 15.*

Theorem 6 (Properties of the Predictive distribution) *Let assumption 4 and 5 hold. Then*

1. $\frac{\partial Var(Y^0)}{\partial \sigma_\varepsilon^2} \geq 0$
2. $\frac{\partial Var(Y^0)}{\partial J} \leq 0$
3. $\frac{\partial Var(Y^0)}{\partial q_i^0} \leq 0$ *where* $q^0 = \left[\int_{x \in Z(\theta^0)\setminus Z(\theta^i)} dF_X^{\theta^0}\right] \times \left[\int_{x \in Z(\theta^i)\setminus Z(\theta^0)} dF_X^{\theta^i}\right]$
4. *If observed data is consistent with the model, then $F^0|_{m,x,\theta} = F_X^{\theta^0}|m,x$*
5. *If the considered model is such that uniquely identifies behavior when $\{\theta^i\}_{i=1}^n$ becomes dense in Θ as $n \to \infty$, then $F^0|_{m,x,\theta} \to F_e^{\theta^0}|m,x,s^0$ where $\{m,s^0\}$ is the limiting behavior.*

where $Y^0 \sim F^0|_{m,x,p}$

Proof. It naturally extends from Theorem 5 ■

The interpretation of these properties is analogous to the ones presented in theorem 5. The measures proposed in section 3.5 naturally extend for the general case.

7 Optimality of Predictive Accuracy as a Measure of the Quality of the Model

Consider the case of a decision maker that must choose a plan of action a to maximize a payoff function g[19] that also depends on a behavioral variable x. Behavior x will be determined after the

[18] This distribution is provided by the researcher, by default it is set to be Uniform.

[19] This function g can be understood as a negative loss function in the Bayesian sense.

choice by the decision maker a, and it may depend on a. For example, a producer may need to set prices to maximize expected profits, but the optimal price depends on the demand function. Given an expected distribution for x, F_x, I assume that the decision maker chooses a to maximize her expected payoff, i.e. $a^* \equiv \arg\max_a \int g(a,x)dF_x$. The decision maker is given with K models of behavior $M^1, M^2, \ldots, M^K$ and believes that behavior is generated by a model i plus some idiosyncratic error process, i.e. $x = m^i + \varepsilon$ with $m^i \in M^i$ and data about past behavior. The decision maker selects one of these models to predict behavior in the relevant economic environment given her optimization problem, she can always opt out of this decision and assume that all feasible alternatives are equally likely. Finally, assume that the decision maker has the technology to infer a predictive distribution, $F^i_{\widehat{x}}$, from a given model, M^i, and past data on behavior, for the moment I abstract of the process to construct such distribution. In this circumstances, what is the model the decision maker prefers? As it would be expected, the decision maker favors, under mild assumptions, unbiased models that deliver more informative predictive distributions.

A predictive distribution being more informative than other depends on the objective function, that is, the particular functional form of $g(\cdot)$ defines the features that make a predictive distribution more informative than other for the considered decision problem. Nevertheless, under weak assumptions, it is possible to identify basic conditions about the first and second moments of the distribution that imply a strict preference order among models. This partial order needs not complete.

Theorem 7 (Optimal model) *Let $M_1, M_2, \ldots, M_K$ be K competing models that, given observed data deliver predictive distributions $F^1_{\widehat{x}}, F^2_{\widehat{x}}, \ldots, F^K_{\widehat{x}}$ respectively. Let $\succ_{DM}$ be the preference relation of the decision maker over models, such that $M^i \succeq_{DM} M^j \Leftrightarrow E_{\widehat{x}}\left(g\left(a^*|F^i_{\widehat{x}}, x^0\right)\right) > E_{\widehat{x}}\left(g\left(a^*|F^j_{\widehat{x}}, x^0\right)\right)$, where $a^*|F^k_{\widehat{x}} \equiv \arg\max_a \int g\left(a, \widehat{x}\right) dF^k_{\widehat{x}}$ for all $k = 1, \ldots, K$ and x^0 is the true behavior in the new economic environment. Moreover, assume that g is twice continuous differentiable and strictly quasiconcave. Then,*

1. *If $Var_{F^i_{\widehat{x}}}\left(\widehat{x}\right) = Var_{F^j_{\widehat{x}}}\left(\widehat{x}\right)$,*

$$\left|E_{F^i_{\widehat{x}}}(\widehat{x} - x^0)\right| \leq \left|E_{F^j_{\widehat{x}}}(\widehat{x} - x^0)\right| \Rightarrow M^i \succeq_{DM} M^j$$

2. *If $\left|E_{F^j_{\widehat{x}}}(\widehat{x} - x^0)\right| = \left|E_{F^i_{\widehat{x}}}(\widehat{x} - x^0)\right|$,*

$$Var_{F^i_{\widehat{x}}}\left(\widehat{x}\right) \leq Var_{F^j_{\widehat{x}}}\left(\widehat{x}\right) \Rightarrow M^i \succeq_{DM} M^j$$

3. *Let $I(M^i) \equiv -\left|AE_{F^i_{\widehat{x}}}(\widehat{x} - x^0) + B\left[\left(E_{F^i_{\widehat{x}}}(\widehat{x} - x^0)\right)^2 + Var_{F^j_{\widehat{x}}}\left(\widehat{x}\right)\right]\right|$, then if $\lim_{n\to\infty} Var_{F^i_{\widehat{x}}} = 0$, where n is the number of observations $I(\cdot)$ represents $\succeq_{DM}$, that is*

$$M^i \succeq M^j \Leftrightarrow I(M^i) \geq I(M^j)$$

with $A \equiv \frac{\partial^2 g}{\partial x \partial a}\left(a_0^*, x^0\right)$, $B \equiv \frac{1}{2}\frac{\partial^3 g}{(\partial x)^2 \partial a}\left(a_0^*, x^0\right)$, *and* $a_0^* \equiv \arg\max_a g(a, x^0)$.

Proof. See Appendix **??** ■

8 Previous literature

This paper proposes the use of the predictive distribution to measure the success of a nonparametric model of utility maximization while allowing for error. To my knowledge, the construction of such distribution is novel in the literature. Varian (1982) uses the nonparametric demand approach to construct forecasts, but considers only the case of data that is perfectly consistent with the model, identifying the set of choices that if they were to be observed in a new budget set would be consistent with the observed (rational) data. My approach extends this analysis by allowing for error in choices and incorporating them in the construction of the predictive distribution. It can be applied to any choice data set, not only those that satisfy GARP. Adams (2013) allows for error, but identifies only a point forecast estimate by applying the "minimum discrimination information principle". The approach developed in this paper identifies instead a predictive distribution, which is both more informative and can be used to measure the quality of the model.

Microeconomic theory has approached the problem of assessing the quality of the model when observed choices are not perfectly rational by proposing goodness of fit measures based on some intuitively appealing moment of the data associated to the monetary cost of the departures from the model. For example, Echenique *et al.* (2011) proposes the money pump index, that considers the total cost of removing violations as a percentage of total expenditure, by adjusting budget sets in a way that violations are not longer feasible. Afriat (1972a) and Varian (1990) propose similar measures. These measures follow the intuition that the severity of the departures from rationality can be measured by the size of the perturbation to income required to make violations infeasible; providing an easily interpretable measure of the fit of the model, but that cannot be interpreted in terms of the behavior that generates the observed data; which prevents the researcher to construct predictions based on such adjustments. The utility maximization model prescribes that choices are the result of decision maker's maximization of her own utility function over the set of feasible choices defined by the observed budgets sets, then adjusting income to remove violations makes impossible to infer any information about behavior or to establish what these adjustments imply in terms of the primitives of the model. Moreover, these adjustments are not the result of a statistical model for choice which impedes the interpretation of different results in terms of their statistical significance. The approach presented in this paper is based on the assumption that behavior is consistent with the model but choices are observed with error, and proposes to assess the quality of the model by the informativeness of the predictive distribution. These measures are easily interpretable and allow for the assessment of the statistical significance of the results.

A common problem that plagues the literature of goodness of fit is the effect of the relative distribution of budget sets in the ability to detect departures from rationality. A data set may pass the rationality test because it was indeed generated by a rational subject or because the distribution of budget sets is such that it is nearly impossible to observe violations. By construction, measures of fit worsen as the number of observations increases even when data was generated by the same individual facing budget regimes that impose equally demanding constraints on data. Moreover, given a same data generating process, these measures punish designs that impose more demanding constraints on data. The literature has approached the problem by studying the "power" of the rationality test; understood as the probability of finding violations given observed economic environments for some alternative behavioral process. For example, Beatty and Crawford (2011) proposes a measure of predictive success that combines fit and power following the axiomatization provided by Selten (1991). This measure consists in assessing the difference between the passing rate and the relative target area where the latter is defined as the relative size of the set of consistent choices with respect to the set of feasible alternatives. In a similar vein, Hoderlein and Stoye (2009) and Andreoni and Harbaugh (2013) tackle this problem by proposing alternative measures of power. The predicting accuracy measures proposed in this paper combine fit and power by sizing the informational content of the predictive distribution; noisier data increases the entropy of the distribution while as the number of observations increases or the constraints imposed by the model become more demanding, the predictive distribution becomes more informative due to the shrinkage of the set of choices that is consistent with previously observed data. In this way, fit and power are not only accounted for, but the quality of the model is assessed by its usefulness to predict behavior, not just explain observed one.

This paper uses a projection technique to identify the systematic component and the error process from data. Others have proposed the use of similar techniques as Varian (1985) and Fleissig and Whitney (2005), but I further impose feasibility constraints. Under the behavioral assumption that choices are rational but observed with error it naturally follows that the rational component must lie in the set of feasible choices; and by definition observed choices are feasible as well. Halevy *et al.* (2014) proposes a projection technique that considers the feasibility constraints but it is based on a parametric assumption over the utility function, that is the data is projected onto the subset of feasible choices that are consistent with the assumed parametric form, inducing misspecification error. The projection technique employed in this paper only relies on the nonparametric constraints imposed by GARP.

Addressing the concerns of interpretation and significance of the results in terms of fit, or combined fit and power, a series of test statistics have been proposed in the literature. Varian (1985), Fleissig and Whitney (2005) and Hjertstrand (2013) propose test statistics for the null of rationality based on the perturbation needed to satisfy the Afriat (1967) set of inequalities that can be equivalently interpreted as additive/multiplicative error in choices. As explained in section 3 I follow a closely related behavioral assumption but I further impose feasibility constraints explicity, the imposition of such constraints naturally follows from the model, and failing on recognizing them alters the effective distribution of any test statistic by omitting the truncation due to nonnegativity constraints on the rational component. In all these cases, the authors argue that under the null the distribution of the proposed test statistics can be bounded by known distributions.

9 Conclusion

Rationality is one of the most prevalent assumptions in economics but empirically is almost surely violated. GARP provides an elegant nonparametric test for rationality, but only one violation and the data would be declared as inconsistent. The aim of this paper is to assess the quality of the model when the data may not pass the deterministic test relying solely on the necessary and sufficient conditions given by GARP. In this paper I develop a framework that combining fit and power gauges the quality of the model by estimating the predictive ability of the model given observed data. The tradeoff between these two features of the model is made conditional on observed data, therefore outperforming other measures of fit prevalent in the literature by considering fit; and other combined measures of fit and power by providing a meaningful tradeoff between these two.

The proposed predictive accuracy measures exhibit, theoretically and empirically, the desired dynamics. Noisier data performs worse and more demanding environments deliver more informative predictive distributions. Moreover, increasing the number of observations allows the researcher to extract more information from data, which results in more accurate predictions.

By combining fit and power forecasting ability measures, are more informative than standard measures in the literature. By relying solely on the axiomatization provided by GARP, without any further parametric/nonparametric assumption on behavior and allowing for individual specific behavior, I avoid misspecification biases common in the recoverability/demand estimation literature. By imposing the constraints implied by the model when projecting the data, I overcome the concerns raised by Lewbel (2001) with respect to the common econometric approaches. Furthermore, I develop intuitive measures based on the information that can be extracted from data given the observed economic environments by imposing the model, which also allows me to develop a simple test statistics based on the significance of this information with respect to an uninformative prior.

Finally, the framework proposed in this paper can be extended to a general class of economic problems, where the model defines a subset of feasible alternatives that are consistent with the constraints imposed by the model, provided that a projection onto the set of consistent choices exists.

References

Adams, A. (2013). Globally rational nonparametric demand forecasts.

Afriat, S. N. (1967). The construction of utility functions from expenditure data. *International Economic Review*, **8** (1), pp. 67–77.

— (1972a). Efficiency estimation of production functions. *International Economic Review*, **13** (3), pp. 568–598.

— (1972b). The theory of international comparisons of real income and prices. In *International Comparisons of Prices and output*, NBER, pp. 11–84.

Andreoni, G. B., James and Harbaugh, W. (2013). The power of revealed preference tests: ex-post evaluation of experimental design. *Working paper*.

Apesteguia, J. and Ballester, M. A. (2013). A measure of rationality and welfare. *Working Papers (Universitat Pompeu Fabra. Departamento de Economía y Empresa)*.

Beatty, T. K. and Crawford, I. A. (2011). How demanding is the revealed preference approach to demand? *The American Economic Review*, **101** (6), 2782–2795.

Becker, G. S. (1962). Irrational behavior and economic theory. *The Journal of Political Economy*, **70** (1), 1–13.

Blundell, R., Browning, M. and Crawford, I. (2008). Best nonparametric bounds on demand responses. *Econometrica*, **76** (6), 1227–1262.

—, Kristensen, D. and Matzkin, R. (2014). Bounding quantile demand functions using revealed preference inequalities. *Journal of Econometrics*, **179** (2), 112–127.

Blundell, R. W., Browning, M. and Crawford, I. A. (2003). Nonparametric engel curves and revealed preference. *Econometrica*, **71** (1), pp. 205–240.

Box, D. N., George E. P. (1987). Empirical model-building and response surfaces. *JOhn Wiley & Sons, New York, NY*.

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, **71** (356), pp. 791–799.

Bronars, S. G. (1987). The power of nonparametric tests of preference maximization. *Econometrica: Journal of the Econometric Society*, pp. 693–698.

Cherchye, L., Crawford, I., De Rock, B. and Vermeulen, F. (2009). The revealed preference approach to demand. *Contributions to Economic Analysis*, **288**, 247–279.

Choi, S., Fisman, R., Gale, D. and Kariv, S. (2007a). Consistency and heterogeneity of individual behavior under uncertainty. *American Economic Review*, **97** (5), 1921–1938.

—, —, Gale, D. M. and Kariv, S. (2007b). Revealing preferences graphically: an old method gets a new tool kit. *The American economic review*, **97** (2), 153–158.

Dean, M. and Martin, D. (2013). Measuring rationality with the minimum cost of revealed preference violations. *Working paper*.

Echenique, F., Lee, S. and Shum, M. (2011). The money pump as a measure of revealed preference violations. *Journal of Political Economy*, **119** (6), 1201–1223.

Famulari, M. (1995). A household-based, nonparametric test of demand theory. *The Review of Economics and Statistics*, pp. 372–382.

Fleissig, A. R. and Whitney, G. A. (2005). Testing for the significance of violations of afriat's inequalities. *Journal of Business & Economic Statistics*, **23** (3), 355–362.

Fostel, A., Scarf, H. E. and Todd, M. J. (2004). Two new proofs of afriat's theorem. *Economic Theory*, **24** (1), 211–219.

Gabaix, X. and Laibson, D. (2008). The seven properties of good models. *The foundations of positive and normative economics. Oxford University Press, Oxford*, pp. 292–299.

Halevy, Y., Persitz, D. and Zrill, L. (2014). *Parametric Recoverability of Preferences*. Tech. rep., Microeconomics. ca Website.

Hjertstrand, P. (2013). *A Simple Method to Account for Measurement Errors in Revealed Preference Tests*. Working Paper Series 990, Research Institute of Industrial Economics.

— and Swofford, J. (2009). A stochastic test of the number of revealed preference violations. *Working paper*.

Hoderlein, S. and Stoye, J. (2009). Revealed preferences in a heterogeneous population. *Review of Economics and Statistics*.

Houthakker, H. S. (1950). Revealed preference and the utility function. *Economica*, **17** (66), pp. 159–174.

Houtman, M. and Maks, J. (1985). Determining all maximal data subsets consistent with revealed preference. *Kwantitatieve methoden*, **19**, 89–104.

Kelley, C. (1999). Detection and remediation of stagnation in the nelder–mead algorithm using a sufficient decrease condition. *SIAM Journal on Optimization*, **10** (1), 43–55.

Kocoska, L. (2012). *A non-parametric approach to the estimations of critical inputs to economic models based on consumption data*. Ph.D. thesis.

Lagarias, J. C., Reeds, J. A., Wright, M. H. and Wright, P. E. (1998). Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on Optimization*, **9** (1), 112–147.

Lewbel, A. (2001). Demand systems with and without errors. *American Economic Review*, pp. 611–618.

Mas-Colell, A. (1978). On revealed preference analysis. *The Review of Economic Studies*, pp. 121–131.

McKinnon, K. I. (1998). Convergence of the nelder–mead simplex method to a nonstationary point. *SIAM Journal on Optimization*, **9** (1), 148–158.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The computer journal*, **7** (4), 308–313.

Pardo, L. (2005). *Statistical inference based on divergence measures*. CRC Press.

Price, C. J., Coope, I. and Byatt, D. (2002). A convergent variant of the nelder–mead algorithm. *Journal of Optimization Theory and Applications*, **113** (1), 5–19.

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, **5** (17), pp. 61–71.

— (1948). Consumption theory in terms of revealed preference. *Economica*, **15** (60), pp. 243–253.

Selten, R. (1991). Properties of a measure of predictive success. *Mathematical Social Sciences*, **21** (2), 153–167.

Varian, H. R. (1982). The nonparametric approach to demand analysis. *Econometrica: Journal of the Econometric Society*, pp. 945–973.

— (1985). Non-parametric analysis of optimizing behavior with measurement error. *Journal of Econometrics*, **30** (1), 445–458.

— (1990). Goodness-of-fit in optimizing models. *Journal of Econometrics*, **46** (1), 125–140.

— (2006). Revealed preference. *Samuelsonian economics and the twenty-first century*, pp. 99–115.

A Proofs of Section 2

.

Claim 1 *If GARP is satisfied then there exists an index $i \in J$ with $a_{ij} \geq 0$ for all $j \in J$*

Proof of Claim 1. (From Fostel *et al.* (2004)) If this were not so, then every row would have an strictly negative entry. Let i be such that $a_{ij} < 0$. Now consider row j and identify a negative entry, say $a_{jk} < 0$. In this fashion construct the sequence $i, j, k, \ldots$ until an index is repeated. This sequence yields a contradiction to GARP. ■

Proof of Proposition 1. If GARP is satisfied then for every chain $\{i, j, k, \ldots, r\} \subset J$, $a_{ij} \leq 0$, $a_{jk} \leq 0$, ...,$a_{ri} \leq 0$ implies that all terms are zero.

From claim 1 we can prove that if GARP is satisfied then there exists a permutation such that $a'_{kj} \geq 0$ for all $j = 1, \ldots, m$. First, define a bijective function $f : J \rightarrow J$ where $f(i)$ corresponds at the row/columns assigned in matrix A' to the row/column i in matrix A. From claim 1 there exists a $i \in J$ such that $a_{ij} \geq 0$ for all $j \in J$. Set $f(i) = J$. Consider now the matrix $A_{(J-1)\times(J-1)}$ from the matrix A eliminating column and row i. Since the data set that contains all the observations but i, also satisfies GARP it must be the case, by claim 1 that there exists a $i' \in J_{-i}$ such that $a_{i'j} \geq 0$ for all $j \in J_{-i}$. Repeat this until the matrix has been reduced to a 1×1 matrix. Finally construct A' by $a'_{ij} = a_{f^{-1}(i),f^{-1}(j)}$. ■

Proof of Proposition 2. Let $a'_{jk} > 0$ for all $k < j \leq J$, then there is no chain of preferences such that $\{i, j, k, \ldots, r\} \subset J$, $a'_{ij} \leq 0$, $a'_{jk} \leq 0$, ...,$a'_{ri} \leq 0$; since $a'_{ij} \leq 0$ implies that $i < j$, $a'_{jk} \leq 0$ implies that $j < k$,....,$a'_{ri} \leq 0$ implies that $r < i$, that is $i < j < k < \cdots < r < i$ which is a contradiction. ■

Proof of Proposition 3. *The condition of lemma 1 is necessary but not sufficient*

Example 1 *Consider $p^1 = (1, 1, 2)$, $x^1 = (1, 0, 0)$, $p^2 = (2, 1, 1)$, $x^2 = (0, 1, 0)$, $p^3 = (1, 2, 1.5)$ and $x^3 = (0, 0, 1)$, then it is the case that $x^1 RP x^2$, $x^2 RP x^3$ and $x^3 RSP x^1$, therefore the data is not consistent with GARP.*

The matrix A' is given by

$$A' = \begin{bmatrix} 0 & .5 & -.5 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

which satisfies the conditions in lemma 1 but the data set generating it is not consistent with GARP.

The condition of lemma 2 is sufficient but not necessary

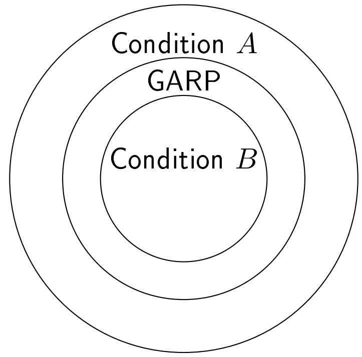


Figure 10: Relationship between GARP and sufficient and necessary conditions

Example 2 *Consider* $p^1 = (1,1,2)$*,* $x^1 = (1,0,0)$*,* $p^2 = (1,1,1)$*,* $x^2 = (0,1,0)$*,* $p^3 = (1,3,1)$ *and* $x^3 = (0,0,2)$*. This data is consistent with GARP, observation 1 and 2 constitutes a violation of WARP [*$x^1(p^1)^T = x^2(p^2)^T = 2$ *and* $x^1(p^2)^T = 1$ *and* $x^2(p^1)^T = 1$ *]*

The matrix A' is given by

$$A' = \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & 0 \\ 3 & 0 & 0 \end{bmatrix}$$

Therefore satisfy the entries in the lower triangle are all positive (not strictly, and therefore not consistent with lemma 2), this data indeed satisfies GARP. ■

The relationship between these conditions is shown in figure 10. From the conditions it is easily seen that the distance between these three sets it is actually zero since $cl(cond_B) = cond_A$ and $cond_B \subset GARP \subset cond_A$.

Remark 1 *Condition A can be seen as the closure of condition B, and therefore the distance between the set of choices that satisfies condition B and the set of choices that satisfies condition A is infinitesimally small.*

B Proof of Section 3.1

Proof of Theorem 2. Let $R \in \mathcal{R}_{\{x^i,p^i\}_{j=1}^{n+1}}$, then since the data is consistent with GARP there exists a preference order that rationalizes the data, that is, there exists a bijective function, $f : I^{n+1} \to I^{n+1}$ that defined the order, where $I^{n+1} = \{1, \ldots, n+1\}$ and such that $f(i) > f(k) \Leftrightarrow x^i R x^k$. Consider now the preference order defined by the injective function $g : I^n \to I^n$ where $I^n = \{1, \ldots, n\}$ given by

$$g(i) = \begin{cases} f(i) & if \quad f(i) < f(n+1) \\ f(i) - 1 & if \quad f(i) > f(n+1) \end{cases}$$

then, given that f defines a complete and transitive preference order so does g, then $R \setminus \{(a,b)|a = x^{n+1}$ or $b = x^{n+1}\}$, then $\mathcal{R}_{\{x^i,p^i\}_{j=1}^{n}} \supseteq \mathcal{R}_{\{x^i,p^i\}_{j=1}^{n+1}}$ ■

Proof of Theorem 4. For any $n \in \mathbb{N}$, there exists $i \in \{1,\ldots n\}$, and $y,z \in \mathbf{B}(x^i,p^i)$ such that $y,z \notin \mathbf{B}(x^j,p^j)$ for all $j \in \{1,\ldots,n\} \setminus \{i\}$ with $y \neq z$. Consider any $M \in \mathbf{M}\left(\tilde{B}^n \setminus \{p^i,x^i\}\right)$ then $\{M,y\},\{M,z\} \in \mathbf{M}\left(\tilde{B}^n\right)$. Now, under assumption 3, there exists a $m > n$ such that $\mathbf{B}(x^i,p^i) \setminus \mathbf{B}(x^m,p^m) \neq \emptyset$, $\mathbf{B}(x^m,p^m) \setminus \mathbf{B}(x^i,p^i) \neq \emptyset$ and $p^m y < p^m z = P^m x^m$. Then $\{M,y,\tilde{x}^m\} \notin \tilde{B}^n \cap \{x^m,p^m\}$ while $\{M,z,\tilde{x}^m\} \in \tilde{B}^n \cup \{x^m,p^m\}$ for any $\tilde{x}^m \in \mathbf{B}(x^m,p^m)$ such that $\{M,\tilde{x}^m\} \in \mathbf{M}\left(\left[\tilde{B}^n \cup \{x^m,p^m\}\right] \setminus \{p^i,x^i\}\right)$. ■

C Projection Procedure

The projection procedure defined in section 3.3 is based on finding the minimal perturbations necessary for the perturbed choices satisfied the condition from lemma **??**. Then the objective is to find a matrix A^ε such that $a^\varepsilon_{kj} = \langle p^*_k, (x^*_j + \varepsilon_j) - (x^*_k + \varepsilon_k)\rangle$, $k,j \in \{1,2,...,J\}$, where ε is the residual and A^ε is reordered according to the matrix A^* where the demand vector includes the residual vector. The constraints placed in the problem are the following

$$\begin{aligned}
\langle p^*_k, \varepsilon^+_j - \varepsilon^-_j\rangle - \langle p^*_k, \varepsilon^+_k - \varepsilon^-_k\rangle &\geq -a^\varepsilon_{kj} &(12)\\
\langle p^*_k, \varepsilon^+_k - \varepsilon^-_k\rangle &= 0 &(13)\\
\varepsilon^+_k - \varepsilon^-_k &\geq -x^*_k &(14)\\
\varepsilon^+_k, \varepsilon^-_k &\geq 0 &(15)\\
\forall \qquad & k = 1,....,J, \text{ and } j = 1,...,k-1 &(16)
\end{aligned}$$

Where $\varepsilon_i = \varepsilon^+_i - \varepsilon^-_i$, and $\varepsilon^+_i - \varepsilon^-_i$ are the positive and negative parts of ε_i. Equation 12 constrains the elements in the lower triangle to be positive, consistent with Lemma 1; equation 13 constrains the perturbed consumption bundle to remain in the budget; while equations 14 and 15 impose non negativity constrains.

The problem, as defined above, depends on the particular order and perturbation, therefore the optimization should be done with respect to finding the optimal ε for each particular re-order and finding the optimal re-order given the minimal residuals that can be achieved for each particular order. Kocoska (2012) proposes the implementation of a process that automates the re-ordering of the data sets while still minimizing the chosen objective value. The pattern search method that is introduced is an adapted version of Nelder and Mead (1965) simplex method which evaluates a function value at a finite number of points and chooses a descendent direction at each step according to this function value; but it has been modified to ensure convergence on the based of a "flexible pattern search".

The Nelder-Mead algorithm has been broadly used because it does not require the knowledge of derivatives and it is easy to use. McKinnon (1998) shows that this method can fail to converge or converge to non-stationary solutions on certain classes of problems. Some limited convergence results have been introduced in the literature, see Lagarias *et al.* (1998) in low dimensions. Price *et al.* (2002) present a convergent variant of the Nelder-Mead algorithm, following Kelley (1999) definition of a sufficient descendent condition. The variant of the Nelder-Mead algorithm followed by Kocoska (2012) is based in Price *et al.* (2002) variant imposing a "flexible pattern search" extending the convergence proof to the case where differentiability nor a Lipschitz property are assumed.

Convergence results and proof can be found in Kocoska (2012), convergence is ensured for a continuously differentiable objective functions that are bounded below and therefore consistent with the maintained assumptions in this paper. There is convergence to a non-zero residual and then convergence to a stationary point follows from strict differentiability of the convex objective functions at non zero points. See theorem 3.4.4 in Kocoska (2012).

D Proofs

D.1 Section 6.2

Proof of Theorem 5. [Proof this theorem properly] Property 1 follows from $\frac{\partial F_{e}^{p^0,x^0}|m,x,v^0}{\partial \sigma_{\varepsilon}^2} > 0$ from assumption 1 and 2 and from equation 2.

Property 2 follows from proposition 4

Property 3 follows from the the fact that the supporting set weakly shrinks as q_0 increases. (prove this)

Property 4 follows from definition 16.

Property 5 follows from theorem 3 and definition 16. ■

D.2 Section 3.5

Proof of Lemma 5. (1) Follows from the properties of probability

(2) If $\widehat{x} = x$ by definition of the residuals of the projection $\widehat{\varepsilon} = 0$, then $F^0|_{m,x,p} = F_X|x \in B(p^0, x^0)$. An the result follows from the definition of $PA_{\alpha}^{p^0,x^0}|_{m,x,p}$

(3)Follows from $\frac{\partial \sigma_{\widehat{x}}}{\partial \sigma_{\varepsilon}} > 0$ and the fact that $\frac{\partial \left(F^0|_{m,x,p}\right)^{(-1)}\left(\frac{\alpha}{2}\right)}{\partial \sigma_{\widehat{x}}} < 0$ and $\frac{\partial \left(F^0|_{m,x,p}\right)^{(-1)}\left(1-\frac{\alpha}{2}\right)}{\partial \sigma_{\widehat{x}}} > 0$ and the properties of the process for ε given by assumption 1.

(4) Follows directly from definition 16.

(5) Follows from theorem 5 and assumptions 1 and 2. ■

D.2.1 Information theoretical measures

Let $F^1 \equiv F^0|_{m,x,p}$ be the predicted distribution of choices based on the model and observed data as given by definition 16 and $F^2 \equiv F_X|x \in B(p^0, x^0)$ be the prior (assumed) distribution of choices. Let f^1 and f^2 be their density functions, and p^1, p^2 their discretized version respectively[20]. Then, the divergence measures considered are defined as follows

Definition 17 (Kullback-Leibler divergence measure) *The Kullback-Leibler divergence measure is defined as,*

$$\begin{aligned} PA^{p^0,x^0}_{info-KL}|_{m,x,p} &= D_{KL}(F^1||F^2) \\ &= \textstyle\sum_{z_i \in B(p^0,x^0)} \ln\left(\frac{p^1_{z_i}}{p^2_{z_i}}\right) p^1_{z_i} \end{aligned} \tag{17}$$

Definition 18 (Hellinger divergence measure) *The Hellinger divergence measure is given by*

$$\begin{aligned} PA^{p^0,x^0}_{info-Hell}|_{m,x,p} &= \tfrac{1}{\sqrt{2}} \left\| \sqrt{F^1} - \sqrt{F^2} \right\|_2 \\ &= \tfrac{1}{\sqrt{2}} \textstyle\sum_{z_i \in B(p^0,x^0)} \left(\sqrt{p^1_{z_i}} - \sqrt{p^2_{z_i}} \right)^2 \end{aligned} \tag{18}$$

Definition 19 (Total variation measure) *The total variation divergence measure is given by,*

$$\begin{aligned} PA^{p^0,x^0}_{info-TV}|_{m,x,p} &= \tfrac{1}{2} \|F^1 - F^2\|_1 \\ &= \tfrac{1}{2} \textstyle\sum_{z_i \in B(p^0,x^0} \left| p^1_{z_i} - p^2_{z_i} \right| \end{aligned} \tag{19}$$

Proposition 7 (Connections from information theory among the measures given by definition 12) *Let F^1, F^2, $PA^{p^0,x^0}_{info-i}|_{m,x,p}$ be given as in definition 12 then,*

1. $PA^{p^0,x^0}_{info-KL}|_{m,x,p} = -H(f^1) - E_{f^1}\left(\ln(f^2)\right)$ *where $H(\cdot)$ is the Shannon entropy measure. Furthermore, if the prior is uniform, $PA^{p^0,x^0}_{info-KL}|_{m,x,p} = \ln n - H(F^1)$ where n corresponds to the number possible values that x may take on $B(p^0, x^0)$.*

2. $PA^{p^0,x^0}_{info-Hell}|_{m,x,p} = \sqrt{1 - BC\left(F^1, F^2\right)}$ *where $BC\left(F^1, F^2\right) = \sum_{x \in X} \sqrt{p^1(x)p^2(x)}$ is the Battacharyya coefficient*

3. $\left(PA^{p^0,x^0}_{info-Hell}|_{m,x,p}\right)^2 \leq PA^{p^0,x^0}_{info-TV}|_{m,x,p} = \sqrt{2} PA^{p^0,x^0}_{info-Hell}|_{m,x,p}$

[20] Consider a partition $\mathcal{P} = \{z_i\}_{i=1,\ldots,n}$ of the set of feasible choices $Z(\theta_{J+1})$ in n intervals. Let $\mathbf{p^1} = (p^1_1, \ldots, p^1_n)$ and $\mathbf{p^2} = (p^2_1, \ldots, p^2_n)$ be the probabilities of the predicted and prior distribution respectively of the intervals z_i, in such a way that $p^j_i = \int_{z_i} dF^j$ for $i = 1, \ldots, n$ and $j = 1, 2$

4. $PA^{p^0,x^0}_{info-TV}|_{m,x,p} \leq PA^{p^0,x^0}_{info-KL}|_{m,x,p}$

Proof of lemma 7. (1) It follows from the definition of Kullback-Leibler discrepancy measure and Shannon entropy.

(2) Follows from the definitions.

(3) Follows from the definition of the 1-norm and 2-norm and the definitions.

(4) Follows from Pinsker's inequality. ■

Maximizing the Kullback-Leibler measure is equivalent to minimizing Shannon entropy, that is the uncertainty in terms of predicting an outcome using the estimated distribution of choices -(1)-. Hellinger discrepancy measure is related with the Battacharyya coefficient as in -(2)-. This coefficient provides a measure of overlapping of the two distributions, where $0 \leq BC \leq 1$ and $BC = 0$ if there is no overlap at all between the two distributions and $BC = 1$ if there is perfect overlap. Finally -(3)- and -(4)- establishes bounds for the total variation measure in terms of the Hellinger and the Kullback-Leibler discrepancy measures.

Proof of Lemma 6. (1) Follows from the properties of these measures.

(2) If $\widehat{x} = x$ ($\widehat{\varepsilon_j} = 0$ for all j), then $Dom(F^1) = V\left(p^0, x^0|m\right)$, follows from definition 16.

Assume now that F_X is uniform. Then $F^1 = F^2|V\left(p^0, x^0|m\right)$. Let $\gamma \equiv \int_{V(p^0,x^0|m)} f^2(x)dx$, then

$$p^1_{z_i} = \begin{cases} \frac{p^2_{z_i}}{\gamma} & for \quad z_i \in V\left(p^0, x^0|m\right) \\ 0 & for \quad z \in B(p^0, x^0)/V\left(p^0, x^0|m\right) \end{cases}$$

Then

$$\begin{aligned} PA^{p^0,x^0}_{info-KL}|_{m,x,p} &= \sum_{z_i \in V(p^0,x^0|m)} \ln\left(\frac{p^1_{z_i}}{p^2_{z_i}}\right) p^1_{z_i} = \sum_{z_i \in V(p^0,x^0|m)} \ln\left(\frac{\frac{p^2_{z_i}}{\gamma}}{p^2_{z_i}}\right) \frac{p^2_{z_i}}{\gamma} \\ &= \sum_{z_i \in V(p^0,x^0|m)} \ln\left(\frac{1}{\gamma}\right) \frac{p^2_{z_i}}{\gamma} = \ln\left(\frac{1}{\gamma}\right) = -\ln\gamma \end{aligned}$$

$$
\begin{aligned}
PA^{p^0,x^0}_{info-Hell}|_{m,x,p} &= \frac{1}{\sqrt{2}} \sum_{z_i \in V(p^0,x^0|m)} \left(\sqrt{p^1_{z_i}} - \sqrt{p^2_{z_i}}\right)^2 \\
&= \frac{1}{\sqrt{2}} \sum_{z_i \in V(p^0,x^0|m)} \left(\sqrt{\frac{p^2_{z_i}}{\gamma}} - \sqrt{p^2_{z_i}}\right)^2 + \frac{1}{\sqrt{2}} \sum_{z_i \in B(p^0,x^0))/V(p^0,x^0|m)} \left(-\sqrt{p^2_{z_i}}\right)^2 \\
&= \frac{1}{\sqrt{2}} \left[\sum_{z_i \in V(p^0,x^0|m)} p^2_{z_i} \left(\frac{1-\sqrt{\gamma}}{\sqrt{\gamma}}\right)^2 + (1-\gamma)\right] \\
&= \frac{1}{\sqrt{2}} \left[\gamma \frac{\left(1-\sqrt{\gamma}\right)^2}{\gamma} + (1-\gamma)\right] \\
&= \frac{1}{\sqrt{2}} \left[(1-\sqrt{\gamma})^2 + (1-\gamma)\right] = \sqrt{2}\left[1-\sqrt{\gamma}\right]
\end{aligned}
$$

$$
\begin{aligned}
PA^{p^0,x^0}_{info-TV}|_{m,x,p} &= \frac{1}{2} \sum_{z_i \in B(p^o,x^0)} \left|p^1_{z_i} - p^2_{z_i}\right| \\
&= \frac{1}{2} \left[\sum_{z_i \in V(p^0,x^0|m)} \left|\frac{p^2_{z_i}}{\gamma} - p^2_{z_i}\right| + \sum_{z_i \in Z(\theta_{J+1})/V(p^0,x^0|m)} \left|-p^2_{z_i}\right|\right] \\
&= \frac{1}{2} \left[\sum_{z_i \in V(p^0,x^0|m)} \left|p^2_{z_i}\left(\frac{1}{\gamma} - 1\right)\right| + \sum_{z_i \in Z(\theta_{J+1})/V(p^0,x^0|m)} p^2_{z_i}\right] \\
&= \frac{1}{2} \left[\left(\frac{1}{\gamma} - 1\right) \sum_{z_i \in V(p^0,x^0|m)} p^2_{z_i} + (1-\gamma)\right] \\
&= \frac{1}{2} \left[\frac{1-\gamma}{\gamma}\gamma + (1-\gamma)\right] = \frac{1}{2}\left[(1-\gamma) + (1-\gamma)\right] = 1-\gamma
\end{aligned}
$$

(3) Follows from properties of the process for ε given by assumption 1 and the definition of the predicted distribution 16

(4) Follows directly from definition 16. ■

E Algorithms

Algorithm 1 (Computation of σ^2_ε assuming $\widehat{x} = m$) *Input: choice data x, observed economic environments $\left\{p^i, x^i\right\}_{i=1}^J$, projection $\widehat{x}$ and σ^2*

Output: vector $S^1 \in \mathbb{R}^M_+$ where M is the number of repetitions for the simulation.

1. *Choose M and set, $m = 1$*

2. *Draw e^m_j from $N\left(0, \sigma^2\right)$ for $j = 1, \ldots, J$, and compute $y^m = \widehat{x} + e^m$*

3. *If $y^m \in \mathbf{B}(p,x)$, set $x^m = y^m$ and go to 4. Otherwise compute x^m from y^m consistently with assumptions 1 and 2*

4. *If $x^m \in \mathbf{M}(\mathbf{B}(p,x))$ then set $S^1_m = 0$ and $m = m+1$. Otherwise compute $ee = x^m - \widehat{x}$, and set $S^1_m = \frac{1}{J} ee' ee$.*

5. *Repeat steps 2-4 M times*

Algorithm 2 (Computation of σ^2_ε by simulating $\mathbf{M}(\mathbf{B}(p,x))$ sequentially) *Input: data on choices x, observed economic environments $\left\{p^i, x^i\right\}_{i=1}^J$, projection $\widehat{x}$ and σ^2*

Output: vector $S^2 \in \mathbb{R}^M$ where M is the number of repetitions for the simulation.

1. *Choose M and set, $m = 1$*

2. *Find a randomize order I^m and set $i = 1$*

3. *take $I^{m(i)}$ and draw $z^m_{I^{m(i)}} \sim U\left[Varian\_Supp\left(\theta_{I^{m(i)}}\right)|\widehat{x}_{-I^{m(i)}}\right]$*

4. *Draw e^m_j from $N\left(0,\sigma^2\right)$ for $j = 1,\ldots,J$, and compute $y^m = z^m_j + e^m$*

5. *If $y^m \in \mathbf{B}(p,x)$, set $x^m = y^m$ and go to 6. Otherwise compute x^m from y^m consistently with assumptions 1 and 2*

6. *If $x^m \in \mathbf{M}(\mathbf{B}(p,x))$ then set $S^2_m = 0$ and $m = m+1$. Otherwise compute $ee = \min\left\{x^m - \widehat{x}, x^m - z^m\right\}$, and set $S^2_m = \frac{1}{J} ee' ee$.*

7. *Repeat steps 3-6 M times*

Algorithm 3 (Computation of σ^2_ε by simulating $\mathbf{M}(\mathbf{B}(p,x))$) *Input: data on choices x, observed economic environments $\left\{p^i, x^i\right\}_{i=1}^J$, projection $\widehat{x}$ and σ^2*

Output: vector $S^3 \in \mathbb{R}^M$ where M is the number of repetitions for the simulation.

1. *Simulate the rational choice set*
   - (a) *Choose N, and set $n = 1$ $i = 1$, and set $JN = \emptyset$*
   - (b) *Choose j at random from the set $JJ = \{1,\ldots,J\} \setminus JN$ and set $I^{n(i)} = j$*
   - (c) *Draw $z^n_{I^{n(j)}} \sim U\left[Varian\_Supp\left(\mathbf{B}(p,x)_{I^{n(i)}}\right)|\widehat{x}_{JJ}\right]$*
   - (d) *Set $JN = JN \cup I^{n(i)}$ and repeat steps 1b-1c until $JJ = \emptyset$*
   - (e) *Set $M^n_{sim}\left(\mathbf{B}(p,x)\right) = z^n$*

2. *Simulate the distribution of the Q statistic*
   - (a) *Choose M and set, $m = 1$*

(b) *Choose* z^m *at random from the set* $M_{sim}\left(\mathbf{B}(p,x)\right)$

(c) *Draw* e_j^m *from* $N\left(0,\sigma^2\right)$ *for* $j=1,\ldots,J$*, and compute* $y^m = z_j^m + e^m$

(d) *If* $y^m \in \mathbf{B}(p,x)$*, set* $x^m = y^m$ *and go to 2e. Otherwise compute* x^m *from* y^m *consistently with assumptions 1 and 2*

(e) *If* $x^m \in \mathbf{M}(\mathbf{B}(p,x))$ *then set* $S_m^3 = 0$ *and* $m = m+1$*. Otherwise compute* $ee = x^m - \widehat{x}$*, and set* $S_m^3 = \frac{1}{J}ee'ee$*.*

(f) *Repeat steps 2b-2e* M *times*

F Tables and Figures: Experimental data from Choi *et al.* (2007a)

F.1 Alternative Goodness of Fit Measures

Blundell *et al.* (2008) proposed that when allowing for a stochastic component when conciliating the data with the revealed preference conditions for utility maximization, the measure of the distance between data and restricted estimators of demand provide a natural formulation for a test statistic for the null of rationality. Alternative, one can construct goodness of fit measures as a measure of the distance of data from the models. I propose to construct: (1) a raw goodness of fit, (2)an adjusted goodness of fit and (3) a weighted goodness of fit. The raw goodness of fit is defined as the coefficient of determination constructed from the residuals of the projection. Formally,

Definition 20 (Raw R^2) *Let* $\widehat{\varepsilon}$ *be defined as the residuals from the projection exercise then*

$$R_l^2 = 1 - \frac{\sum_{j=1}^{J}\left(\widehat{\varepsilon}_j^l - \overline{\widehat{\varepsilon}}^l\right)^2}{\sum_{j=1}^{J}\left(x_j^l - \overline{x}^l\right)^2} \tag{20}$$

As other measures of fit proposed in the literature, the empirical fit of the rationality model is tightly related to the ex-ante probability of detecting violations to the model given the observed economic environments and not only noisy, and the former is particularly significant in small samples. Them I propose to adjust the measure of fit the ex ante probability of detecting violations to the model given observed economic environments. These measure is closely related to Beatty and Crawford (2011), though the considered measure here is the product of the two factors instead of the difference. Formally,

Definition 21 (Adjusted measure of goodness of fit) *Define the adjusted measure of goodness of fit as*

$$\overline{R}^2 \equiv \textit{Adjusted Goodness of fit} \equiv \textit{Goodness of fit} \times Pr[\textit{Violation model}] \tag{21}$$

where Goodness of fit is defined as in 21 and $Pr[$*Violation model*$]$ *is defined as the mean of the ex-ante probability of detecting a violation to the model given observed economic environments.*

Finally, the weighted measure of fit is a measure of fit where each of observation is weighted by how unlikely (ex-ante) is to observed such behavior given observed economic environments; that is,

Definition 22 (Adjusted measure of goodness of fit - Weighted R^2) *If there exists a j such that $p_j \neq 0$, then the weighted measure of goodness of fit is given by*

$$R^2_{l,weighted} = 1 - \frac{\sum_{j=1}^{J} w_j \left(\widehat{\varepsilon}^l_j - \overline{\widehat{\varepsilon}}^l\right)^2}{\sum_{j=1}^{J} w_j \left(x^l_j - \overline{x}^l\right)^2} \tag{22}$$

otherwise it is equal to zero.

The weights are given by

Definition 23 (Weights) *Consider the weight function w_j such that*

$$\widetilde{w}_j = \begin{cases} p_j & if \quad \widehat{\varepsilon}_j = 0 \\ 1 - p_j & if \quad \widehat{\varepsilon}_j \neq 0 \end{cases} \tag{23}$$

and $w_j = \frac{\widetilde{w}_j}{\sum_j^J \widetilde{w}}$ with $p_j \equiv Pr\left(violation_{j|-j} | x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_J\right)$

F.2 Data

Table 6: Summary statistics for the proposed measures and other standard measures from the literature applied to the experimental data from Choi *et al.* (2007a).

Measure	Mean	Median	Std. dev.	Min	Max
Rational?	0.1975	0.0000	0.4006	0.0000	1
Afriat	0.9568	0.9810	0.0642	0.6860	1.0000
Varian	0.8898	0.9440	0.1568	0.2290	1.0000
HM index	47.1235	48	3.2418	29	50
# violations WARP	4.8889	4.0000	5.3852	0	27
# violations GARP	43.2099	6.0000	104.2718	0	559
Mean ex-ante prob viol	0.1142	0.1077	0.0670	0.0059	0.3449
Raw R^2	0.9906	0.9994	0.0214	0.8597	1.0000
Adjusted R^2	0.1072	0.1023	0.0640	0.0009	0.2806
Weighted R^2	0.9906	0.9994	0.0207	0.8720	1.0000
Considering all the observed Budget set					
PA CI (95%)	0.4935	0.4918	0.0546	0.3607	0.6199
PA KL	0.3072	0.2603	0.1590	0.1729	1.1049
PA Hellinger	0.2962	0.2879	0.0866	0.1654	0.6006
PA TV	0.3940	0.3834	0.0615	0.2775	0.5916

Raw R^2, adjusted R^2 and weighted R^2 are defined as in 20, 21 and 22 respectively

Table 7: Correlations - Data for 81 subjects, Choi et al (2007).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) PA CI (95%)	1											
(2) PA KL	0.0898	1										
(3) PA Hellinger	0.7303*	0.4117*	1									
(4) PA TV	0.7867*	0.4795*	0.9628*	1								
(5) Raw R^2	0.1663	-0.7304*	-0.1044	-0.1839	1							
(6) Adjusted R^2	0.1043	-0.0702	0.1168	0.0588	0.1777	1						
(7) Weighted R^2	0.1859	-0.7259*	-0.0830	-0.1672	0.9974*	0.1665	1					
(8) Ex ante prob viol	0.0591	0.1606	0.0850	0.1314	-0.1511	0.0997	-0.1533	1				
(9) # viol WARP	-0.3589*	0.4585*	-0.1613	-0.0472	-0.7100*	-0.1073	-0.7354*	0.1092	1			
(10)# viol GARP	-0.2015	0.5547*	0.0380	0.1433	-0.7321*	-0.0742	-0.7455*	0.0318	0.7797*	1		
(11)Afriat	0.1898	-0.6885*	-0.0667	-0.1725	0.8635*	0.1202	0.8662*	-0.1015	-0.8126*	-0.8642*	1	
(12)Varian	0.2217*	-0.5969*	-0.0209	-0.1041	0.8360*	0.1153	0.8430*	-0.0546	-0.7984*	-0.8552*	0.9323*	1
(13)HM	0.2788*	-0.4536*	0.0621	-0.0607	0.4642*	0.1572	0.4897*	-0.1319	-0.7825*	-0.7215*	0.6205*	0.5531*

* significance $p < 0.05$

F.3 Performance Predictive Ability Measures and Comparison to the Literature

Table 8: Predictive Ability CI

	Forecasting Ability: Confidence Interval (95%)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
R^2	1.390***		1.458***	1.434***	1.358**	1.411***	1.339**
	(0.503)		(0.508)	(0.523)	(0.533)	(0.521)	(0.531)
Ex ante mean P		0.041	0.095	0.131*	0.140*	0.127	0.136*
		(0.093)	(0.068)	(0.075)	(0.074)	(0.077)	(0.076)
rational					0.014		0.013
					(0.017)		(0.017)
Adj. R^2						0.062	0.057
						(0.091)	(0.092)
Constant	-0.886*	0.488***	-0.965*	-0.947*	-0.875	-0.930*	-0.863
	(0.500)	(0.011)	(0.504)	(0.519)	(0.529)	(0.515)	(0.525)
R-squared	0.150	-0.010	0.154	0.160	0.158	0.153	0.151
N	80	80	80	73	73	73	73

Raw R^2 and adjusted R^2 are defined as in 20 and 21. Significance * p<0.10, ** p<0.05, *** p<0.01. The sample has been restricted to those observations that show an Afriat index>.7, Varian index>.3 and raw R^2>.85, which results in drops subject 105 (603 in Choi's nomination). Columns (4)-(7) are restricted to those observations for what the mean ex ante probability of detecting a violation to rationality is above 0.025.

Table 9: Predictive Ability CI Comparison Afriat

	Predictive Ability: Confidence Interval (95%)			
	(1)	(2)	(3)	(4)
Afriat	0.299*	-0.074	0.306*	-0.093
	(0.130)	(0.200)	(0.129)	(0.203)
R^2		1.615		1.745
		(0.858)		(0.889)
Ex ante mean P			0.064	0.099
			(0.073)	(0.073)
Constant	0.205	-1.039	0.191	-1.161
	(0.126)	(0.706)	(0.122)	(0.733)
R-squared	0.087	0.141	0.082	0.145
N	80	80	80	80

Raw R^2 and adjusted R^2 are defined as in 20 and 21. Significance * p<0.10, ** p<0.05, *** p<0.01. The sample has been restricted to those observations that show an Afriat index>.7, Varian index>.3 and raw R^2>.85, which results in drops subject 105 (603 in Choi's nomination).

Table 10: Predictive Ability KL

	Predictive Ability: Kullback Liebler						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
R^2	-4.830***	259.107*	-4.731***	257.991*	304.715**	300.567**	299.538**
	(1.084)	(135.140)	(1.081)	(134.291)	(151.421)	(146.653)	(146.512)
$(R^2)^2$		-135.812*		-135.191*	-159.384**	-155.888**	-155.222**
		(69.716)		(69.261)	(78.149)	(75.369)	(75.215)
Ex ante mean P			0.137	0.124	0.114		0.194
			(0.155)	(0.145)	(0.149)		(0.178)
Adj R^2					0.277		
					(0.228)		
Afriat						-0.850**	-0.900**
						(0.376)	(0.389)
Constant	5.089***	-123.055*	4.976***	-122.572*	-145.139*	-143.606**	-143.217**
	(1.084)	(65.431)	(1.080)	(65.044)	(73.304)	(71.132)	(71.121)
R-squared	0.317	0.402	0.313	0.399	0.416	0.445	0.445
N	80	80	80	80	73	73	73

Raw R^2 and adjusted R^2 are defined as in 20 and 21. Significance * p<0.10, ** p<0.05, *** p<0.01. The sample has been restricted to those observations that show an Afriat index>.7, Varian index>.3 and raw R^2>.85, which results in drops subject 105 (603 in Choi's nomination). Columns (5)-(7) are restricted to those observations for what the mean ex ante probability of detecting a violation to rationality is above 0.025.

Table 11: Predictive Ability Hellinger

	Predictive Ability: Hellinger					
	(1)	(2)	(3)	(4)	(5)	(6)
R^2	1.179**		1.149**	1.081**	1.645*	1.972**
	(0.508)		(0.532)	(0.513)	(0.829)	(0.858)
Ex ante mean P		0.179	0.221*	0.208		0.238*
		(0.129)	(0.130)	(0.126)		(0.133)
Adj R^2				0.184		
				(0.183)		
Afriat					-0.203	-0.264
					(0.229)	(0.219)
Constant	-0.877*	0.268***	-0.877	-0.828	-1.148*	-1.442**
	(0.504)	(0.015)	(0.532)	(0.516)	(0.663)	(0.708)
R-squared	0.040	0.008	0.051	0.060	0.025	0.049
N	80	73	73	73	73	73

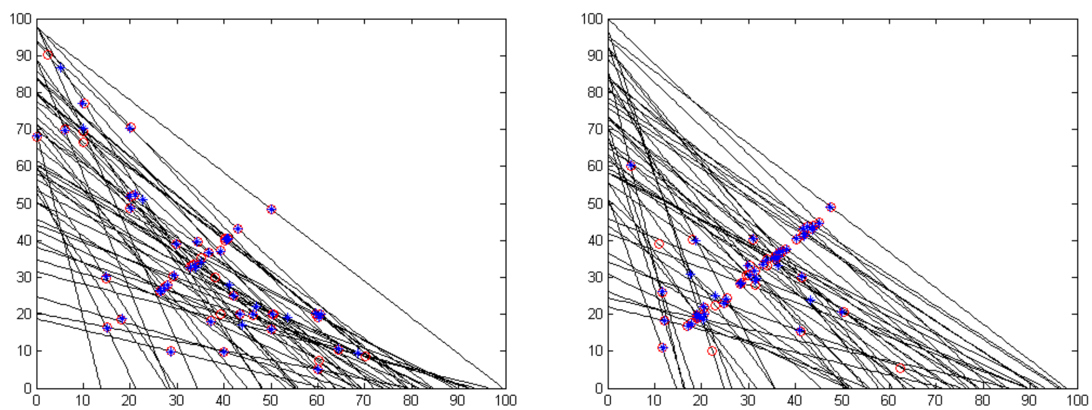
Raw R^2 and adjusted R^2 are defined as in 20 and 21. Significance * p<0.10, ** p<0.05, *** p<0.01. The sample has been restricted to those observations that show an Afriat index>.7, Varian index>.3 and raw R^2>.85, which results in drops subject 105 (603 in Choi's nomination). Columns (2)-(6) are restricted to those observations for what the mean ex ante probability of detecting a violation to rationality is above 0.025.

Table 12: Predictive Ability Total Variation

	Predictive Ability: Total Variation					
	(1)	(2)	(3)	(4)	(5)	(6)
$R^2 2$	0.355		0.358	0.331	0.983	1.242*
	(0.400)		(0.417)	(0.415)	(0.679)	(0.692)
Ex ante mean P		0.158*	0.171*	0.166*		0.190**
		(0.086)	(0.090)	(0.089)		(0.092)
Adj R^2				0.072		
				(0.117)		
Afriat					-0.235	-0.284
					(0.183)	(0.174)
Constant	0.040	0.371***	0.014	0.034	-0.359	-0.593
	(0.397)	(0.012)	(0.417)	(0.416)	(0.541)	(0.568)
R-squared	-0.004	0.018	0.014	0.007	-0.005	0.026
N	80	73	73	73	73	73

Raw R^2 and adjusted R^2 are defined as in 20 and 21. Significance * p<0.10, ** p<0.05, *** p<0.01. The sample has been restricted to those observations that show an Afriat index>.7, Varian index>.3 and raw R^2>.85, which results in drops subject 105 (603 in Choi's nomination). Columns (2)-(6) are restricted to those observations for what the mean ex ante probability of detecting a violation to rationality is above 0.025.

Figure 11: Comparison of measures for two irrational subjects



(a) Subject 209

(b) Subject 606

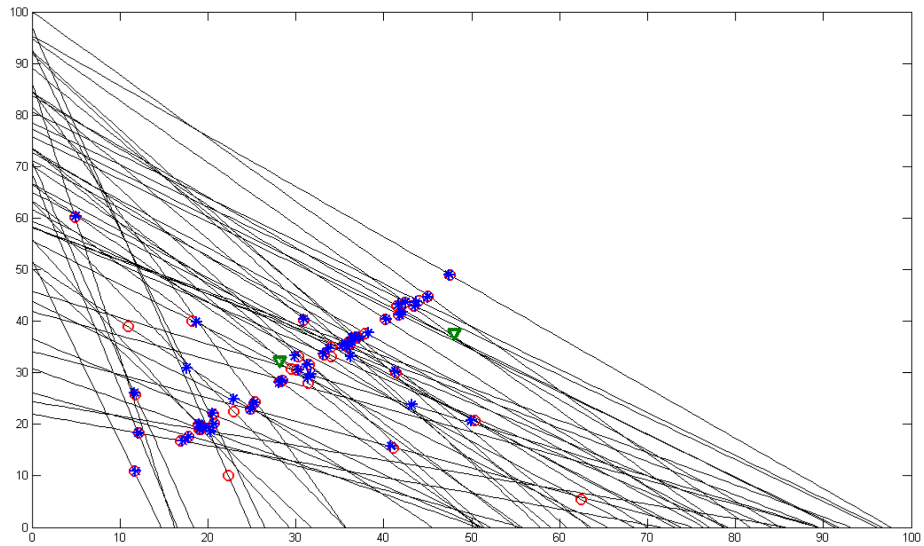
F.4 Changes in the number of Observations

Table 13: Changes in Predictive Ability Measures

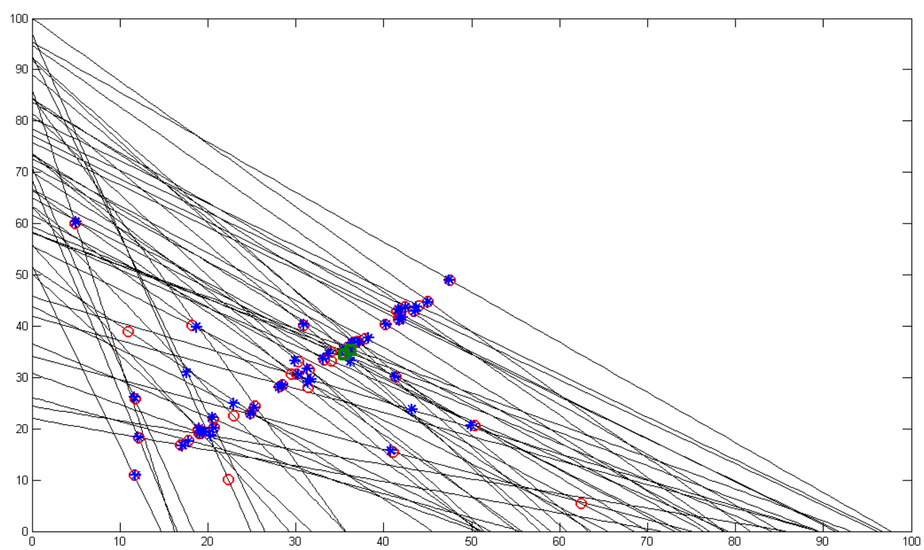
	PA CI		PA KL			PA Hellinger		FA TV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$R^2_{full} - R^2_{half}$	0.892*	0.945**	-3.711***	-3.675***		0.261	0.313		-0.009
	(0.474)	(0.438)	(1.317)	(1.298)		(0.853)	(0.814)		(0.475)
$p_{full} - p_{half}$		0.217*		0.148			0.203	0.186	0.186
		(0.129)		(0.130)			(0.204)	(0.140)	(0.138)
p					0.452*				
					(0.252)				
Constant	0.102***	0.095***	0.095***	0.090***	0.064**	0.084***	0.077***	0.085***	0.085***
	(0.010)	(0.010)	(0.017)	(0.016)	(0.025)	(0.015)	(0.017)	(0.012)	(0.012)
R-squared	0.059	0.085	0.364	0.361	0.021	-0.011	-0.009	0.012	-0.002
N	78	78	78	78	78	77	77	77	77

Raw R^2 and adjusted R^2 are defined as in 20 and 21. p refers to the ex ante mean probability to detect violations to the model given observed economic environments. Significance * p<0.10, ** p<0.05, *** p<0.01.

F.5 The effect of parametric assumptions on demand

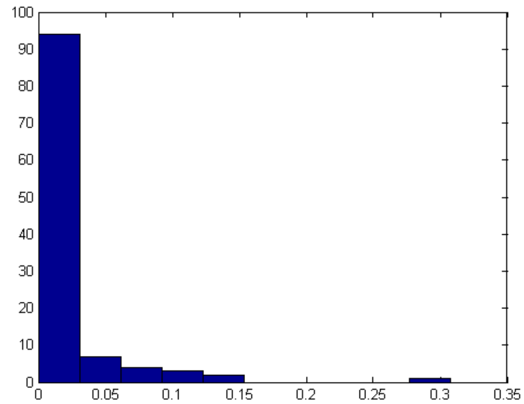


(a) Cobb-Douglas preferences

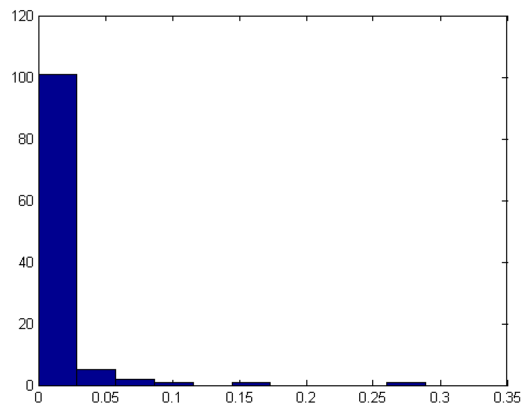


(b) Leontief preferences

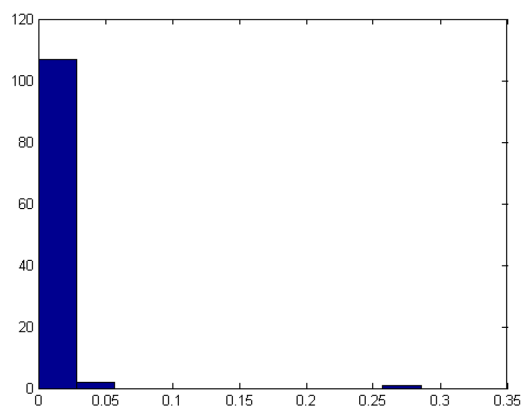
Figure 12: Data, projection based solely on GARP constraints and parametric projection for subject 606. Labels: ○ data, (*) GARP, △ Cobb-Douglas preferences, and □ Leontief preferences.



(a) Assumption Cobb Douglas Preferences



(b) Assumption Leontief Preferences



(c) Assumption Linear Preferences

Figure 13: The effect of parametric assumption on the SSR, with respect to the necessary assumptions for GARP

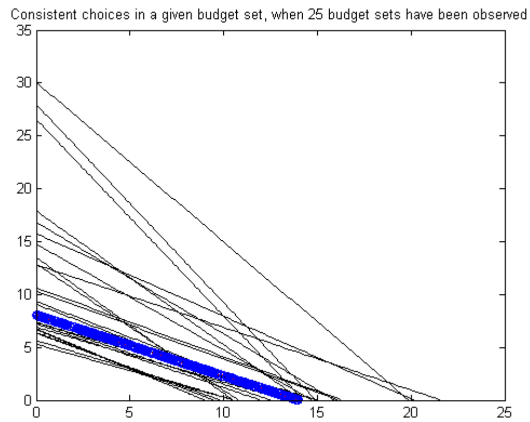
G Monte Carlo Simulations

5.2

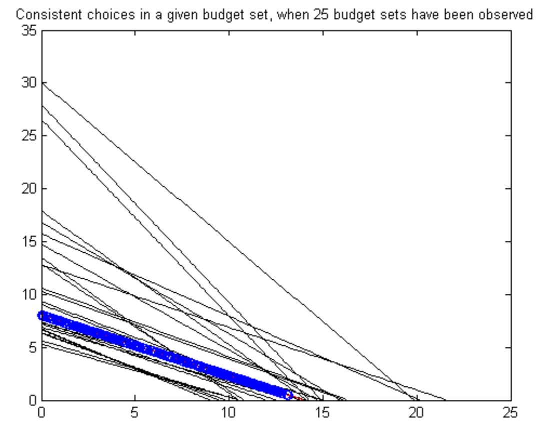
G.1 Changes in the number of observations

	$J = 10$	$J = 25$	$J = 50$	$J = 100$
Rational?	0.9192	0.4596	0.0303	0.0000
	(0.2732)	(0.4996)	(0.1719)	(0.0000)
R^2	0.9999	0.9990	0.9966	0.9852
	(0.0005)	(0.0057)	(0.0065)	(0.0123)
Adj R^2	0.0720	0.0725	0.0745	0.0732
	(0.0567)	(0.0539)	(0.0503)	(0.0436)
Adj R^2 **BC**	0.0820	0.0812	0.0805	0.0588
	(0.0738)	(0.0708)	(0.0638)	(0.0543)
Weighted R^2	0.9999	0.9990	0.9966	0.9853
	(0.0005)	(0.0057)	(0.0065)	(0.0122)
PA CI (95%)	0.2071	0.3353	0.4385	0.5542
	(0.1720)	(0.2082)	(0.2157)	(0.2061)
PA KL	0.0669	0.1155	0.1886	0.4042
	(0.0654)	(0.0824)	(0.1307)	(0.2542)
PA Hellinger	0.0934	0.1951	0.2863	0.3893
	(0.1298)	(0.1833)	(0.2235)	(0.2210)
PA TV	0.1493	0.2724	0.37296	0.4892
	(0.1620)	(0.2019)	(0.2145)	(0.2036)

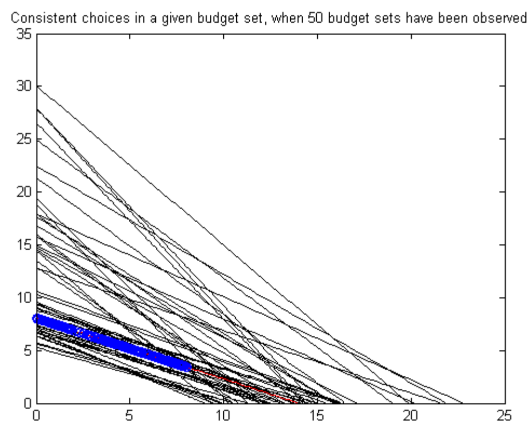
Table 14: Data for 198 simulations, where choices where generated as $x = m+\varepsilon$, where m is the optimal demand from a utility function of the form $u(x, y) = \sqrt{xy}$ and $\varepsilon \sim N(0, 1)$. Four different sample sizes are considered, $J \in \{10, 25, 50, 100\}$.



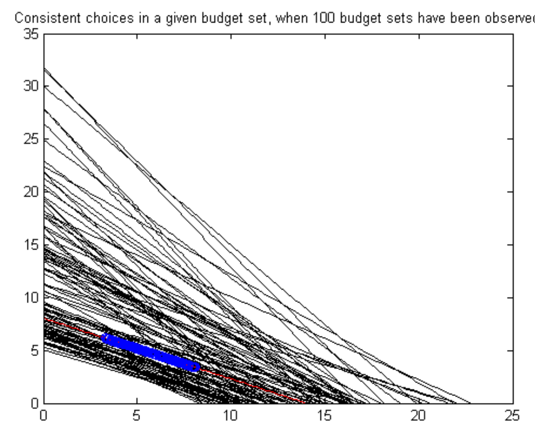
(a) 10 choices



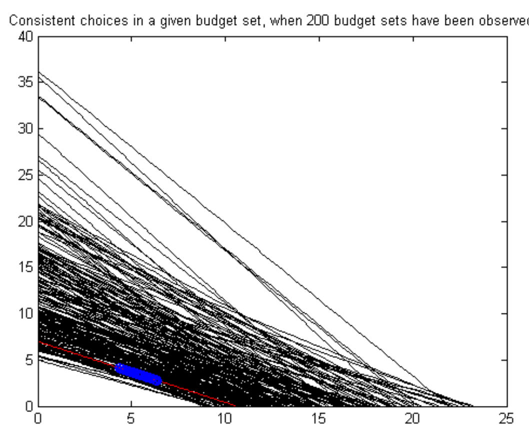
(b) 25 choices



(c) 50 choices



(d) 100 choices



(e) 200 choices

(f) 1000 choices

Figure 14: Convergence of Identified set as the number of observations increases. The considered choices were generated from a DM with a CD with $\alpha = \frac{1}{2}$ and without noise. In particular, 100%, 91.60%, 61.00%, 35.80% and 16.80% of the feasible alternatives in the reference budget sets are consistent with the DM's choices for the first 10,25,50,100 and 200 budget sets.

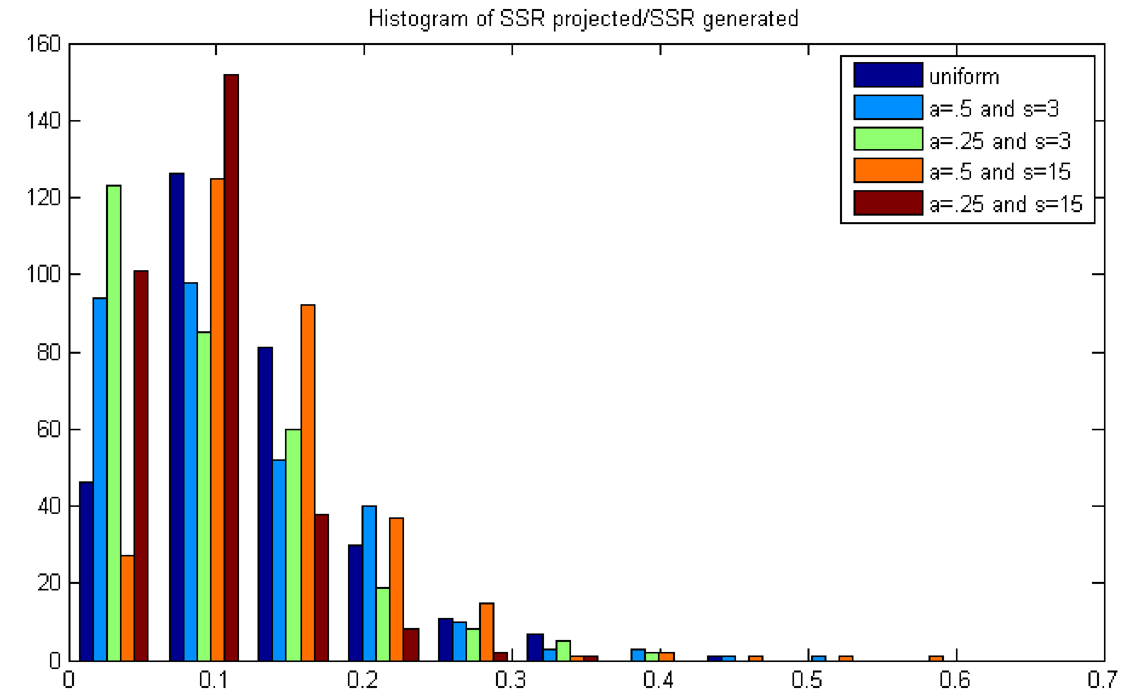


Figure 15: Histogram of ratio of generated to projected SSR for simulated process with $\sigma = 3, 15$ and baseline CD vs uniform alternative

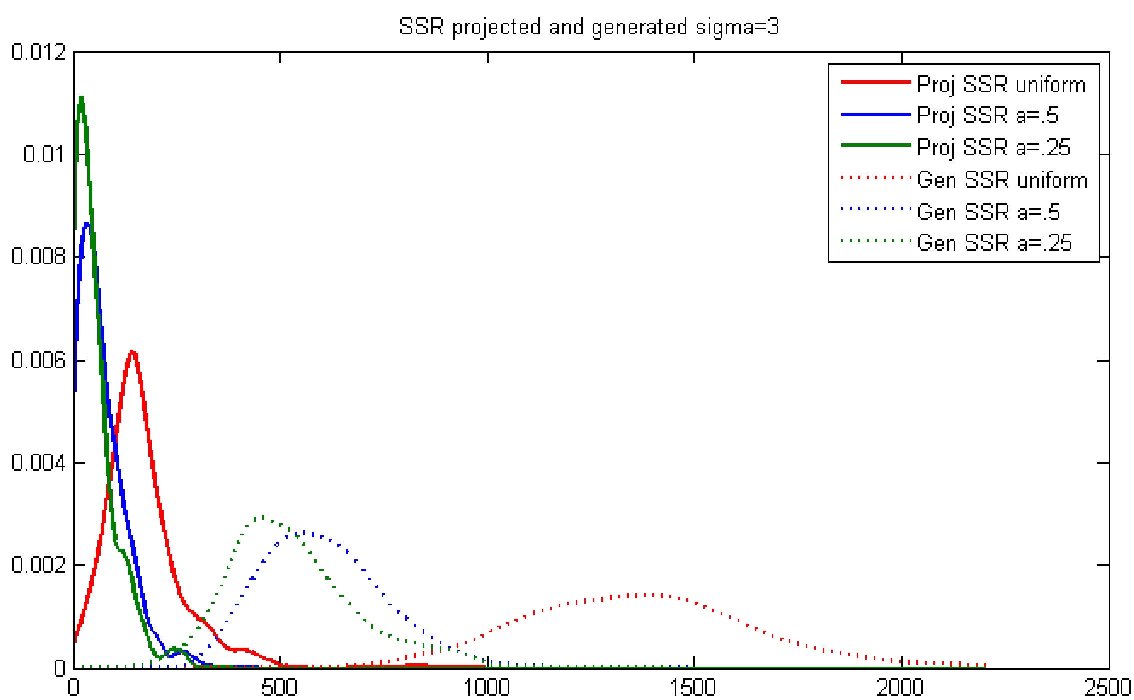
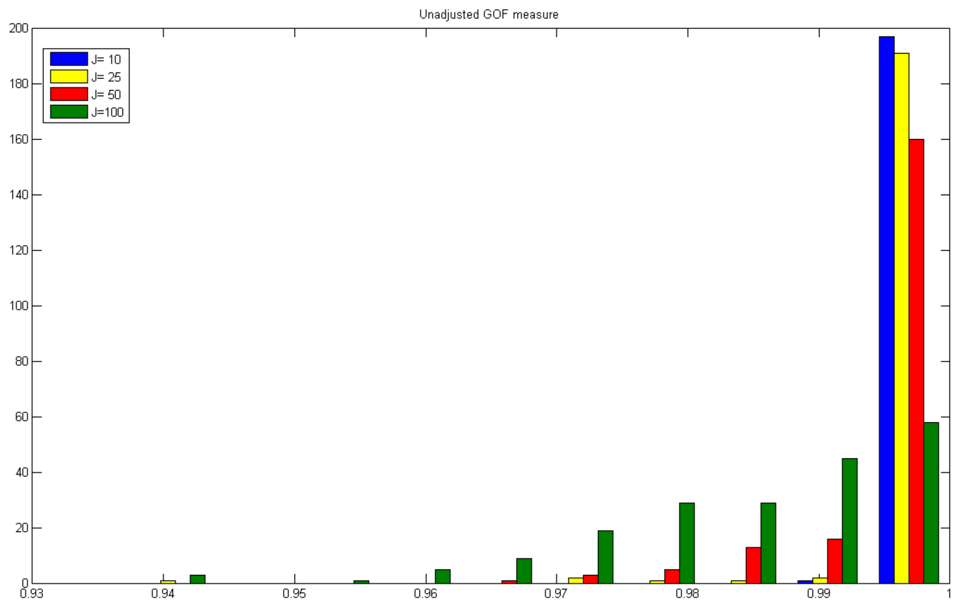
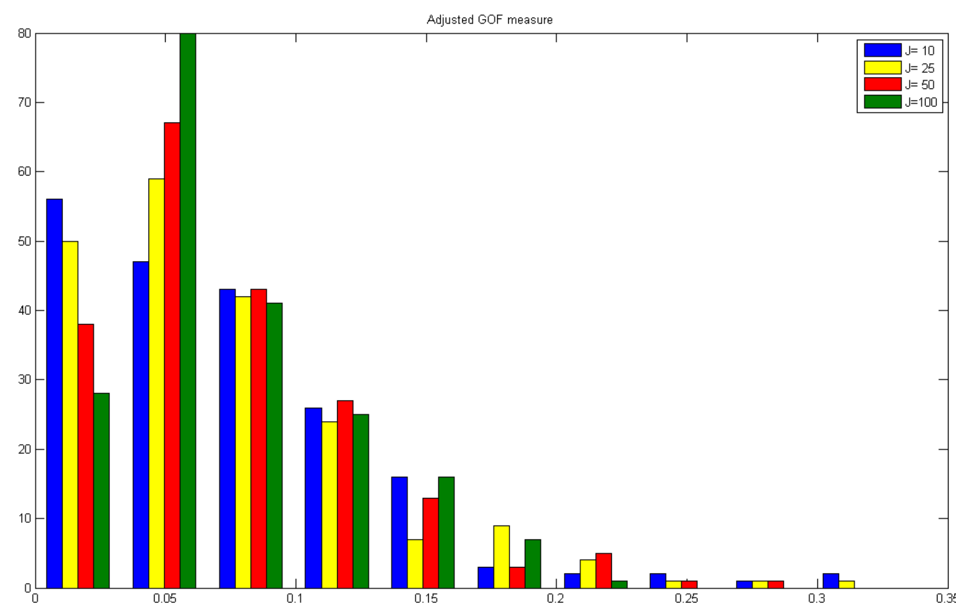


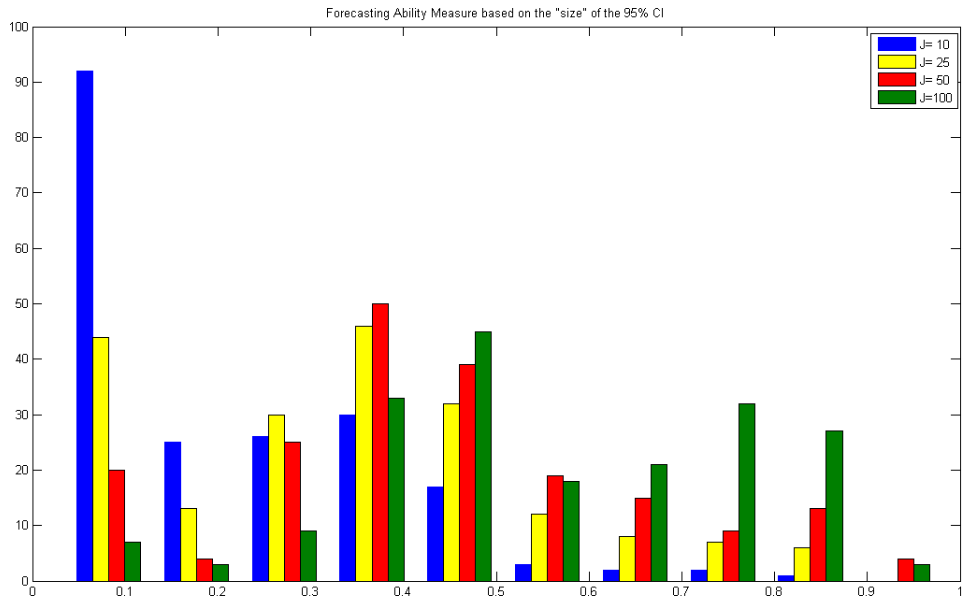
Figure 16: Histogram of generated and projected SSR for simulated process with $\sigma = 3$ and baseline CD vs uniform alternative



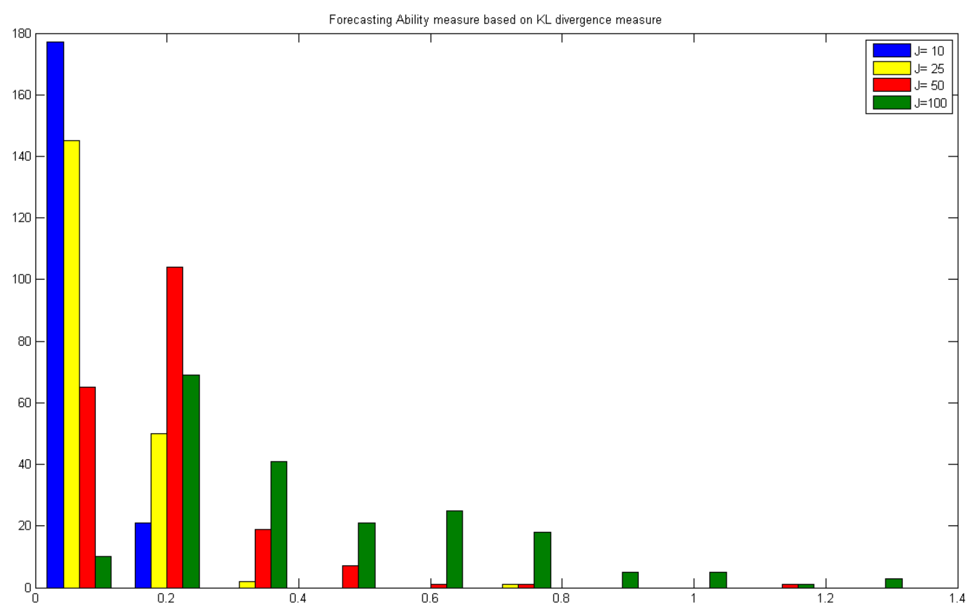
(a) Unadjusted GOF measure



(b) Adjusted GOF measure



(a) PA based on "size" confidence intervals



(b) PA based on KL divergence measures

Figure 18: Empirical distribution of the data for GOF and Predictive accuracy measures for different sample sizes.

G.2 Changes in the DGP

	$\sigma_1 = .5$	$\sigma_2 = 1$	$\sigma_3 = 2$	$\sigma_4 = 5$	$\sigma_5 = 10$
Rational?	0.6350	0.1350	0.0051	0.0000	0.0000
	(0.4826)	(0.3426)	(0.0712)	(0.0000)	(0.0000)
R^2	0.9999	0.9977	0.9697	0.8793	0.7988
	(0.0014)	(0.0072)	(0.0326)	(0.0768)	(0.1171)
Adj R^2	0.0742	0.0719	0.0629	0.0590	0.0600
	(0.0546)	(0.0538)	(0.0418)	(0.0406)	(0.0371)
Adj R^2 **BC**	0.0742	0.0698	0.0346	-0.0531	-0.1247
	(0.0545)	(0.0540)	(0.0544)	(0.0842)	(0.1172)
Weighted R^2	0.9999	0.9976	0.9698	0.8800	0.8023
	(0.0015)	(0.0072)	(0.0328)	(0.0756)	(0.1117)
PA CI (95%)	0.3992	0.4060	0.4123	0.2440	0.2263
	(0.1966)	(0.1913)	(0.2150)	(0.1994)	(0.1950)
PA KL	0.1327	0.1528	0.2863	0.3810	0.6134
	(0.0610)	(0.0765)	(0.2390)	(0.4004)	(0.5661)
PA Hellinger	0.2440	0.2479	0.2574	0.1989	0.2639
	(0.1879)	(0.1769)	(0.2082)	(0.2003)	(0.2347)
PA TV	0.3315	0.3389	0.3613	0.3381	0.4380
	(0.1917)	(0.1858)	(0.2157)	(0.2124)	(0.2023)
Convergent Observations	200	200	197	98	37

Table 15: Data for 200 simulations, where choices where generated as $x = m+\varepsilon$, where m is the optimal demand from a utility function of the form $u(x, y) = \sqrt{xy}$ and $\varepsilon \sim N(0, \sigma_i^2)$, where the five different assumptions correspond to $\sigma_i \in \Sigma = \{.5, 1, 2, 5, 10\}$. Standard deviation in ().Results shown for all the convergent simulations in each case

	$\sigma_1 = .5$	$\sigma_2 = 1$	$\sigma_3 = 2$	$\sigma_4 = 5$
Rational?	0.6350	0.1350	0.0051	0.0000
	(0.4826)	(0.3426)	(0.0712)	(0.0000)
R^2	0.9999	0.9977	0.9697	0.8793
	(0.0014)	(0.0072)	(0.0326)	(0.0768)
Adj R^2	0.0742	0.0719	0.0629	0.0590
	(0.0546)	(0.0538)	(0.0418)	(0.0406)
Adj R^2 **BC**	0.0742	0.0698	0.0346	-0.0531
	(0.0545)	(0.0540)	(0.0544)	(0.0842)
Weighted R^2	0.9999	0.9976	0.9698	0.8800
	(0.0015)	(0.0072)	(0.0328)	(0.0756)
PA CI (95%)	0.4043	0.4099	0.4033	0.2440
	(0.2026)	(0.2003)	(0.2634)	(0.1994)
PA KL	0.1347	0.1530	0.2836	0.3810
	(0.0629)	(0.0736)	(0.2897)	(0.4004)
PA Hellinger	0.2498	0.2525	0.2547	0.1989
	(0.1981)	(0.1877)	(0.2553)	(0.2003)
PA TV	0.3363	0.3418	0.3539	0.3381
	(0.1980)	(0.1946)	(0.2636)	(0.2124)

Table 16: Data for 200 simulations, where choices where generated as $x = m+\varepsilon$, where m is the optimal demand from a utility function of the form $u(x, y) = \sqrt{xy}$ and $\varepsilon \sim N(0, \sigma_i^2)$, where the five different assumptions correspond to $\sigma_i \in \Sigma = \{.5, 1, 2, 5, 10\}$. Standard deviation in ().Results shown for all the simulations that are convergent in all four assumptions over $\sigma \in \{\sigma_1, \sigma_2, \sigma_3, \sigma_4\}$

H Proofs of Section ??

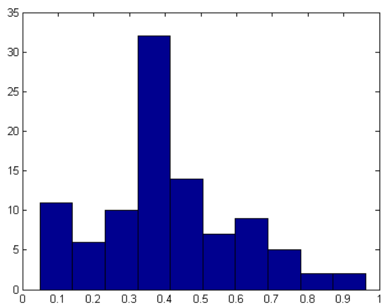
Proof of Theorem 7. *Proof of 1* Define $a^*|F^i_{\widehat{x}} \equiv \arg\max_a \int g(a, \widehat{x})\, dF^i_{\widehat{x}}$ and $a^*(\widehat{x}) \equiv \arg\max_a g(a, \widehat{x})$. Consider now a Taylor expansion of $g(a, \widehat{x})$ around $\widehat{x} = x^0$

$$g(a, \widehat{x}) = g\left(a, x^0\right) + \left(\widehat{x} - x^0\right) \frac{\partial g}{\partial x}\left(a, x^0\right) + \frac{1}{2}\left(\widehat{x} - x^0\right)^2 \frac{\partial^2 g}{(\partial x)^2}\left(a, x^0\right) + o\left(|\widehat{x} - x^0|^2\right)$$

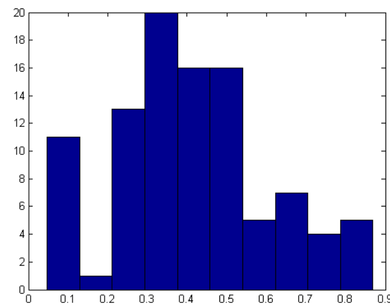
$$\begin{aligned}
a^*|F^i_{\widehat{x}} &\equiv \arg\max_a \int g(a, \widehat{x})\, dF^i_{\widehat{x}} \\
&= \arg\max_a \int \left[g\left(a, x^0\right) + \left(\widehat{x} - x^0\right) \frac{\partial g}{\partial x}\left(a, x^0\right) + \frac{1}{2}\left(\widehat{x} - x^0\right)^2 \frac{\partial^2 g}{(\partial x)^2}\left(a, x^0\right) + o\left(|\widehat{x} - x^0|^2\right)\right] dF^i_{\widehat{x}} \\
&\cong \arg\max_a \left[g\left(a, x^0\right) + \frac{\partial g}{\partial x}\left(a, x^0\right) \int \left(\widehat{x} - x^0\right) dF^i_{\widehat{x}} + \frac{1}{2} \frac{\partial^2 g}{(\partial x)^2}\left(a, x^0\right) \int \left(\widehat{x} - x^0\right)^2 dF^i_{\widehat{x}}\right]
\end{aligned}$$

Then the first order condition is given by

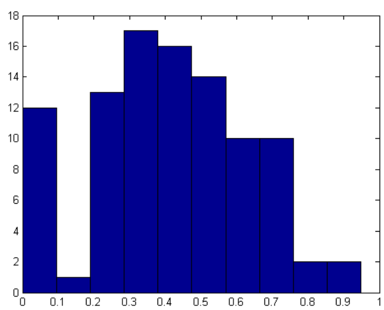
$$\frac{\partial g}{\partial a}\left(a, x^0\right) + \frac{\partial^2 g}{\partial x \partial a}\left(a, x^0\right) \int \left(\widehat{x} - x^0\right) dF^i_{\widehat{x}} + \frac{1}{2} \frac{\partial^3 g}{(\partial x)^2 \partial a}\left(a, x^0\right) \int \left(\widehat{x} - x^0\right)^2 dF^i_{\widehat{x}} = 0$$



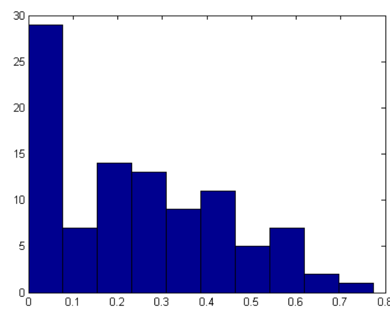
(a) Histogram PA for $\sigma = .5$



(b) Histogram PA for $\sigma = 1$



(c) Histogram PA for $\sigma = 2$



(d) Histogram PA for $\sigma = 5$

Figure 19: Histogram for the Predictive accuracy measure as σ changes, 98 simulations with $J = 40$

Which differs from the first order condition for the case of certainty $a^*(\widehat{x}) \equiv \arg\max_a g(a, \widehat{x})$ by

$$\Delta FOC_{x^0 - F^i_{\widehat{x}}} = -\left[\frac{\partial^2 g}{\partial x \partial a}(a, x^0)\int\left(\widehat{x} - x^0\right)dF^i_{\widehat{x}} + \frac{1}{2}\frac{\partial^3 g}{(\partial x)^2 \partial a}(a, x^0)\int\left(\widehat{x} - x^0\right)^2 dF^i_{\widehat{x}}\right]$$

Then given $Var_{F^i_{\widehat{x}}}(\widehat{x}) = Var_{F^j_{\widehat{x}}}(\widehat{x})$

$$\Delta FOC_{x^0 - F^i_{\widehat{x}}} = -\frac{\partial^2 g}{\partial x \partial a}(a, x^0)\int\left(\widehat{x} - x^0\right)dF^i_{\widehat{x}} = -\frac{\partial^2 g}{\partial x \partial a}(a, x^0)\, E_{F^i_{\widehat{x}}}\left(\widehat{x} - x^0\right)$$

Since $\frac{\partial^2 g}{\partial x \partial a}(a, x^0) \neq 0$ and does not depend on $\widehat{x}$, as bigger the biased on the prediction, the bigger the perturbation to the FOC and, given the strict concavity assumptions on g, the lower the realized payoff.

Proof of 2 It follows from strict concavity and Jensen's inequality.

Proof of 3 From 1 and given assumptions about the partial derivatives of $g(\cdot,\cdot)$, preferences among models of behavior can be established in terms of $\left|a^*\left(x^0\right) - a^*|F^i_{\widehat{x}}\right|$, that is, a lower is the gap between optimal choices given the model and the optimal given the best model (certainty about behavior). Moreover, given the features of g, deviations on the first order condition directly translate on changes in the optimal a and therefore on the ex-post utility,

$$M^i \succeq_{DM} M^j \Leftrightarrow \left|a^*\left(x^0\right) - a^*|F^i_{\widehat{x}}\right| \leq \left|a^*\left(x^0\right) - a^*|F^j_{\widehat{x}}\right| \Leftrightarrow \left|\Delta FOC_{x^0 - F^i_{\widehat{x}}}\right| \leq \left|\Delta FOC_{x^0 - F^j_{\widehat{x}}}\right|$$

■