

Hjertstrand, Per

**Working Paper**

## A simple method to account for measurement errors in revealed preference tests

IFN Working Paper, No. 990

**Provided in Cooperation with:**

Research Institute of Industrial Economics (IFN), Stockholm

*Suggested Citation:* Hjertstrand, Per (2013) : A simple method to account for measurement errors in revealed preference tests, IFN Working Paper, No. 990, Research Institute of Industrial Economics (IFN), Stockholm

This Version is available at:

<https://hdl.handle.net/10419/95647>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

IFN Working Paper No. 990, 2013

# **A Simple Method to Account for Measurement Errors in Revealed Preference Tests**

Per Hjertstrand

# A SIMPLE METHOD TO ACCOUNT FOR MEASUREMENT ERRORS IN REVEALED PREFERENCE TESTS

Per Hjertstrand\*

Research Institute of Industrial Economics (IFN)

Stockholm, Sweden

November 2013

## Abstract

Revealed preference tests are widely used in empirical applications of consumer rationality. These are static tests, and consequently, lack ability to handle measurement errors in the data. This paper extends and generalizes existing procedures that account for measurement errors in revealed preference tests. In particular, it introduces a very efficient method to implement these procedures, which make them operational for large data sets. The paper illustrates the new method for both classical and Berkson measurement errors models.

**Keywords:** Berkson measurement errors; Classical measurement errors; GARP; Revealed preference.

**JEL Classification:** C43; D12.

## 1 Introduction

This paper focus on testing whether mismeasured data is consistent with utility maximization behavior in a revealed preference framework. More precisely, assuming that data on prices and quantities for a set of goods and assets contains measurement errors, the purpose is to provide models and methods to test the null hypothesis that the ‘true’ data (without errors) satisfies utility maximizing behavior, or equivalently, satisfies certain revealed preference axioms. The main contribution is twofold: First, I introduce a simple and very efficient algorithm to implement the revealed preference based procedure for error-contaminated data proposed by Fleissig and Whitney (2005). This algorithm makes it possible to apply revealed preference methods to large scaled data sets (with errors), which are becoming increasingly available in empirical macroeconomics and applied economics in general.<sup>1</sup>

As a second main contribution, I modify Fleissig and Whitney’s (2005) procedure to make it compatible with classical measurement error models. The standard approach to account for measurement errors in revealed preference tests has been to use Berkson error models, where the ‘true’ variable of interest is

---

\*Financial support from the Jan Wallander and Tom Hedelius Foundation (research grant W2009-0079:1) and the Marianne and Marcus Wallenberg Foundation is gratefully acknowledged. I also thank Barry Jones for generously providing the data for the empirical application. Correspondence to: Per Hjertstrand, Research Institute of Industrial Economics (IFN), P.O. Box 55665, SE-102 15 Stockholm, Sweden. E-mail: [Per.Hjertstrand@ifn.se](mailto:Per.Hjertstrand@ifn.se).

<sup>1</sup>Examples of other relevant applications besides empirical macroeconomics where the use of revealed preference tests to analyze large data sets has recently been advocated includes large scale scanner data (Echenique, Lee, and Shum, 2011) and cross-sectional data gathered for a large number of households (Crawford and Pendakur, 2012).

predicted (or caused) by the observed variable and a random error.<sup>2</sup> However, these models have been labeled inappropriate to describe many economic data sets (Chen, Hong and Nekipelov, 2011).<sup>3</sup> Instead, classical measurement error models, where the observed variable is being predicted by the ‘true’ variable and a random error, is widely considered the most appropriate specification to incorporate measurement errors in economic applications (Chen et al., 2011; Hausman, 2001).

Revealed preference analysis provides necessary and sufficient conditions for a data set to be rationalized by a well-behaved utility function. As such, it provides a natural starting point for empirically analyzing dynamic general equilibrium models in macroeconomics since these models are most often based on the assumption that a ‘representative’ consumer maximizes utility over consumption goods and monetary assets. In addition, revealed preference analysis is implicitly a key ingredient in constructing monetary aggregates. This follows from that utility maximization is a necessary condition for monetary aggregates to exist (Barnett, 1980). This, in turn, has motivated numerous studies to use revealed preference analysis to test whether aggregates satisfy this necessary condition; See for examples Swofford and Whitney (1986, 1987, 1994) and Fisher and Fleissig (1997), and more recently Jones, Dutkowsky and Elger (2005), Elger, Jones, Edgerton and Binner (2008), and Jha and Longjam (2006).<sup>4</sup>

From an applied perspective, revealed preference based tests are attractive since they are non-parametric in the sense of not having to stipulate any parametric form for the utility function. However, they are by definition deterministic, and consequently, fail to add any stochastic element to the analysis. A particularly important such element when, for example, analyzing monetary and consumption data is measurement errors (Barnett, Diewert and Zellner, 2009; Belongia, 1996; Fixler, 2009). For this reason, much effort has gone into extending the standard (deterministic) revealed preference tests to make them applicable when there are errors present in the data. Varian (1985), Epstein and Yatchew (1985), de Peretti (2005) and Cherchye, Demuynck, De Rock and Hjertstrand (2012) are a few examples of such test-procedures. However, these procedures share the problem of being computationally burdensome for large or even medium sized data sets. For example, Varian’s (1985) and Epstein and Yatchew’s (1985) procedures are based on solving non-linear programming problems (with non-linear constraints) which may become very computationally burdensome even for moderate sized problems (See Jones and de Peretti, 2005, for a detailed discussion).

As discussed above, Fleissig and Whitney (2005) proposed another revealed preference procedure for dealing with measurement errors in the data. This procedure is based on calculating a test statistic by minimizing the maximal slack term required for the data to satisfy revealed preference, which is, in a second step, compared to a critical value obtained from the empirical distribution of the measurement errors. Jones and Edgerton (2009) suggest calculating Fleissig and Whitney’s (2005) test statistic by solving a non-linear constrained optimization problem. Similar to the procedures mentioned above, it therefore suffers from the problem of being computationally burdensome for medium or large sized data sets. Although it is possible to calculate the test statistic using a binary search algorithm, each step in the binary search consists of solving a linear program (LP) and even if computationally simpler than solving the non-linear problem, solving a LP may still become computationally burdensome for large sized data sets.

---

<sup>2</sup>See, for example, Varian (1985), Epstein and Yatchew (1985), Jones and de Peretti (2005), Elger and Jones (2008) and Hjertstrand (2007).

<sup>3</sup>Berkson models are mostly of interest in the statistics, biology and medical literature; See Carroll, Ruppert, and Stefanski (1995) for motivations and explanations of various Berkson-error models.

<sup>4</sup>Monetary aggregates need to be weakly separable from all other goods in order to qualify as (theoretically) proper aggregates (Barnett, 1980). Revealed preference analysis provides a very convenient way of testing for weak separability. Consequently, empirical studies aimed at finding proper monetary aggregates most often employ revealed preference tests for weak separability. A necessary condition for these tests is that the data satisfies revealed preference.

This paper proposes a very efficient and simple algorithm to calculate Fleissig and Whitney's (2005) test statistic. The new algorithm is especially suitable for the large scaled data sets typically encountered in empirical macroeconomics. The algorithm computes the test statistic from a modified version of the Generalized Axiom of Revealed Preference (GARP). As such, it is based on computing transitive closures of matrices for which there exists very efficient (polynomial-time) methods (Varian, 1982). In fact, results from Monte Carlo experiments (available upon request) show that the algorithm is solvable for several hundreds of observations within reasonable time. Moreover, the algorithm is guaranteed to find a global solution to the test statistic, and perhaps most importantly, it does not require using any optimization software (linear or non-linear) as the case is in Jones and Edgerton (2009).

In this paper, I also extend Fleissig and Whitney's (2005) procedure to classical measurement errors. Given the importance of accounting for measurement errors, this extension allows for analyzing data with errors using models that are widely recognized as the most appropriate ones in empirical economic modelling. As such, the extension to classical measurement error models and the introduction of the new algorithm provides a unified framework for analyzing data with errors within a revealed preference framework.

I demonstrate the practical usefulness of the models and methods by applying them to monetary data from Jones and de Peretti (2005). In this application, I investigate the computational properties of the new algorithm. As explained above, I found the algorithm to be very fast in practice. I also compare and contrast classical measurement error models with Berkson-type models. Rather surprisingly, I found that classical models yields similar results as Berkson models, which would suggest that accounting for measurement errors in revealed preference analysis is robust to the employed error-model. Finally, I compare the different models with respect to the error distribution, and find that the models produce similar results for normally and uniformly distributed errors.

The remainder of the paper is organized as follows: Section 2 briefly recapitulates the concept of revealed preference. Section 3 introduces measurement errors, and describes the new algorithm. Section 4 discusses implementation issues while Section 5 shows how the procedure can accommodate various measurement error models. Section 6 contains the empirical application and Section 7 gives some concluding remarks and discussion.

## 2 Revealed preference

Suppose a consumer chooses from  $K$  goods observed at  $N$  time periods, with index set  $T = \{1, \dots, N\}$ . Let  $\mathbf{x}_t = (x_{1t}, \dots, x_{Kt}) \in \mathbb{R}_+^K$  denote the observed quantity-vector at time  $t \in T$  with corresponding price-vector  $\mathbf{p}_t = (p_{1t}, \dots, p_{Kt}) \in \mathbb{R}_{++}^K$ . Let  $m_t \in \mathbb{R}_{++}$  denote the wealth of the consumer, and define the budget set as  $B_t = B(\mathbf{p}_t, m_t) = \{\mathbf{x} \in \mathbb{R}_+^K : \mathbf{p}_t \cdot \mathbf{x} \leq m_t\}$ . In the standard utility maximizing model, choices are usually taken to be exhaustive which means that budget balancedness holds, i.e.,  $\mathbf{p} \cdot \mathbf{x} = m$ . Throughout the paper, I will refer to the list  $\mathbb{T} = \{\mathbf{p}_t, \mathbf{x}_t, m_t\}_{t \in T}$  as 'the data', but drop  $m$  from this list whenever convenient.

Let us now recall the following concepts and definitions from Varian (1982): we say that the data  $\mathbb{T} = \{\mathbf{p}_t, \mathbf{x}_t, m_t\}_{t \in T}$  is rationalizable if there exists a well-behaved (i.e. continuous, concave and strictly increasing) utility function  $U$  such that the observed consumption bundles solves the utility maximizing problem, i.e.,

$$\{\mathbf{x}_t\}_{t \in T} \text{ solves } \max_{\mathbf{x} \in B_t} U(\mathbf{x}).$$

Consider next the definition of the Generalized Axiom of Revealed Preference (GARP).

**Definition 1 (Generalized Axiom of Revealed Preference, Varian, 1982)** *Consider the data  $\mathbb{T} = \{\mathbf{p}_t, \mathbf{x}_t, m_t\}_{t \in T}$ . We say that:*

- $\mathbf{x}_t$  is directly revealed preferred to  $\mathbf{x}_s$  written  $\mathbf{x}_t R^D \mathbf{x}_s$  if  $m_t \geq \mathbf{p}_t \cdot \mathbf{x}_s$ .
- $\mathbf{x}_t$  is revealed preferred to  $\mathbf{x}_s$  written  $\mathbf{x}_t R \mathbf{x}_s$  if there exists a sequence of observations  $(t, u, v, \dots, w, s) \in T$  such that  $\mathbf{x}_t R^D \mathbf{x}_u$ ,  $\mathbf{x}_u R^D \mathbf{x}_v$ , ...,  $\mathbf{x}_w R^D \mathbf{x}_s$ .
- $\mathbb{T}$  satisfies the Generalized Axiom of Revealed Preference (*GARP*) if  $\mathbf{x}_t R \mathbf{x}_s$  implies  $m_s \leq \mathbf{p}_s \cdot \mathbf{x}_t$ .

In words, *GARP* states that it cannot be that bundle  $\mathbf{x}_t$  is preferred over bundle  $\mathbf{x}_s$  while at the same time the costs for bundle  $\mathbf{x}_t$  at prices  $\mathbf{p}_s$  is strictly less than the costs for bundle  $\mathbf{x}_s$ . Using these concepts and definitions, Varian (1982), based on Afriat (1967), derived necessary and sufficient conditions for the data  $\mathbb{T}$  to be rationalized by a well-behaved utility function.

**Theorem 1 (Afriat’s Theorem, Varian, 1982)** Consider the data  $\mathbb{T} = \{\mathbf{p}_t, \mathbf{x}_t, m_t\}_{t \in T}$ . The following statements are equivalent:

- (i) There exist a continuous, strictly increasing and concave utility function rationalizing  $\mathbb{T}$ .
- (ii)  $\mathbb{T}$  satisfies *GARP*.
- (iii) There exist utility indices  $U_t$  and marginal utility indices  $\lambda_t > 0$  such that the following (Afriat) inequalities hold (for all  $s, t \in T$ ):

$$U_s - U_t \leq \lambda_t (\mathbf{p}_t \cdot \mathbf{x}_s - m_t). \quad (\text{AI})$$

The standard formulation of Afriat’s theorem uses that budget balancedness holds (i.e., sets  $m_t = \mathbf{p}_t \cdot \mathbf{x}_t$  for all  $t \in T$ ). For now, I will keep a separate notation for  $m$  and  $\mathbf{p} \cdot \mathbf{x}$  because the measurement error procedure described in the next section introduces a slack term to allow for violations of the Afriat inequalities. This slack is additive to total expenditure and is basically the minimal perturbation of total expenditure such that the Afriat inequalities hold; I then define  $m$  as the minimally perturbed total expenditure.

Afriat’s theorem presents two different methods for testing whether the data can be rationalized by a well-behaved utility function. The first method implements *GARP* given in condition (ii). This consists in a first step of constructing the  $R^D$  relation. In the second step, one calculates the transitive closure of the relation  $R^D$  (using, for example, Warshall’s (1962) algorithm). The final third step consists of verifying whether  $m_s \leq \mathbf{p}_s \cdot \mathbf{x}_t$  holds whenever  $\mathbf{x}_t R \mathbf{x}_s$ . If this is the case, then  $\mathbb{T}$  satisfies *GARP*, and consequently, the data can be rationalized by a well-behaved utility function. *GARP* can be efficiently implemented (i.e., in polynomial time) using standard statistical and mathematical software, and for that reason has become the most popular test-method in empirical applications of consumer rationality.<sup>5</sup> The second method uses linear programming (LP) techniques to check whether there exists a solution (in the unknowns  $U_t$  and  $\lambda_t$ ) to the Afriat inequalities (AI); See, for example, Diewert (1973). Although this method can be used to find a solution to the Afriat’s inequalities in polynomial time, it is increasingly computationally burdensome as the number of constraints in the LP problem grows quadratically with the number of observations (the number of constraints in the LP problem is  $T^2 - T$ ). Thus, even if computationally efficient for small data sets, this method may become practically infeasible for large data sets.

As discussed in the Introduction, a problem with any one of these methods concerns their inability to incorporate stochastic elements in the analysis. In particular, test-methods based on conditions (ii) and (iii) are inherently deterministic since they are unable to account for any randomness in the data such as measurement errors or optimization errors. This paper focus on measurement errors and introduces next a simple algorithm to implement the measurement error procedure proposed by Fleissig and Whitney (2005).

---

<sup>5</sup>See Varian (2006) for a recent overview.

### 3 Allowing for measurement errors

Suppose the quantity data is measured with random errors collected in the  $K$ -dimensional vector  $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{Kt}) \in \mathbb{R}^K$ . For illustration, I begin by assuming a classical additive measurement error model,

$$\mathbf{x}_t = \mathbf{q}_t + \varepsilon_t, \quad (\text{CA})$$

where  $\{\mathbf{q}_t\}_{t \in T}$  is the 'true' (unobserved) quantity-vector. Below, I show how the classical additive error structure can be modified to a classical multiplicative error structure, and to additive and multiplicative Berkson error structures.

Our purpose is to test the following hypothesis:

$$\begin{aligned} H_0 : & \text{The 'true' data } \{\mathbf{p}_t, \mathbf{q}_t\}_{t \in T} \text{ satisfy GARP} \\ H_A : & \text{The 'true' data } \{\mathbf{p}_t, \mathbf{q}_t\}_{t \in T} \text{ violate GARP.} \end{aligned} \quad (\text{HYP})$$

Here,  $H_0$  corresponds to that there exists a continuous, strictly increasing and concave utility function rationalizing the 'true' (unobserved and without errors) data  $\{\mathbf{p}_t, \mathbf{q}_t\}_{t \in T}$ , while  $H_A$  corresponds to that there do not exist any utility function rationalizing  $\{\mathbf{p}_t, \mathbf{q}_t\}_{t \in T}$ . To test this hypothesis, Fleissig and Whitney (2005, henceforth referred to as FW) proposed a procedure which consists of adding a slack term to the Afriat inequalities (AI) to allow for violations of these. A test statistic is constructed by calculating the minimal slack required for the observed data  $\mathbb{T} = \{\mathbf{p}_t, \mathbf{x}_t\}_{t \in T}$  to satisfy the Afriat inequalities. The hypothesis (HYP) is then evaluated by comparing the test statistic to a critical value obtained from the empirical distribution of the errors.

**The test statistic.** Jones and Edgerton (2009) suggested calculating FW's test statistic, denoted  $F$ , by solving the following optimization problem:

$$\begin{aligned} \min_{\{U_t, \lambda_t, F\}_{t \in T}} \quad & F \quad s.t. & (\text{op\_AI}_F) \\ U_s - U_t - \lambda_t (\mathbf{p}_t \cdot \mathbf{x}_s - \mathbf{p}_t \cdot \mathbf{x}_t) & \leq \lambda_t F, \\ \lambda_t & > 0, \\ F & \geq 0, \end{aligned}$$

for all  $s, t \in T$ . The problem (op\_AI<sub>F</sub>) contains quadratic (non-linear) constraints, which makes it non-trivial. As such, one alternative is to solve (op\_AI<sub>F</sub>) using optimization software that is able to handle quadratic constraints. However, a computationally simpler solution can be obtained by noting that if (op\_AI<sub>F</sub>) has a feasible solution for a specific value of  $F$  then it also has a solution for all values  $F' \geq F$ . This monotonicity condition implies that one can find a solution to  $F$  by applying a binary search algorithm. In practice, this consists of iterating upon  $F$  and check whether there exists a solution to the Afriat inequalities for a given  $F$  by solving a LP problem in each iteration; I discuss the details within a similar context in the next section. However, as pointed out in the previous section, although these LP problems can be solved in polynomial time, they may become inefficient for large scaled data sets, and as a result make the entire procedure difficult to implement in practice.

**A new procedure to calculate the test statistic  $F$ .** To remedy the difficulty of calculating FW's test statistic, the remainder of this section introduces an alternative procedure to calculate this statistic. Motivated by the equivalence between the Afriat inequalities (AI) and **GARP** in Afriat's theorem, the new procedure replaces the non-linear constraints in (op\_AI<sub>F</sub>) with a **GARP**-like condition. As such, this new procedure inherits the advantages of **GARP**. Most importantly from a practical point of view, it does not require using software packages for solving optimization problems (linear or non-linear) as the case is

for  $(\text{op\_AI}_F)$ . For that reason, it is computationally much more efficient than solving  $(\text{op\_AI}_F)$ , and is practically operational for large scaled data sets. But before providing a formal argument for the new procedure, we consider the following condition.

**Definition 2** ( $\text{GARP}_F$ ) *Consider the data  $\mathbb{T} = \{\mathbf{p}_t, \mathbf{x}_t\}_{t \in T}$  and the scalar  $F \geq 0$ . We say:*

- $\mathbf{x}_t R_F^D \mathbf{x}_s$  if  $\mathbf{p}_t \cdot (\mathbf{x}_t - \mathbf{x}_s) \geq F$ .
- $\mathbf{x}_t R_F \mathbf{x}_s$  if there exists a sequence of observations  $(t, u, v, \dots, w, s) \in T$  such that  $\mathbf{x}_t R_F^D \mathbf{x}_u, \mathbf{x}_u R_F^D \mathbf{x}_v, \dots, \mathbf{x}_w R_F^D \mathbf{x}_s$ .
- that  $\mathbb{T}$  satisfies  $\text{GARP}_F$  if  $\mathbf{x}_t R_F \mathbf{x}_s$  implies  $\mathbf{p}_s \cdot (\mathbf{x}_s - \mathbf{x}_t) \leq F$ .

The next theorem proves that  $\text{GARP}_F$  is equivalent to the non-linear constraints in  $(\text{op\_AI}_F)$ .

**Theorem 2** *Consider the data  $\mathbb{T} = \{\mathbf{p}_t, \mathbf{x}_t\}_{t \in T}$  and the scalar  $F \geq 0$ . The following statements are equivalent:*

(i\*) *There exist numbers  $U_t$  and  $\lambda_t > 0$  such that the following inequalities hold (for all  $s, t \in T$ ):*

$$U_s - U_t - \lambda_t (\mathbf{p}_t \cdot \mathbf{x}_s - \mathbf{p}_t \cdot \mathbf{x}_t) \leq \lambda_t F. \quad (\text{AI}_F)$$

(ii\*)  $\mathbb{T}$  satisfies  $\text{GARP}_F$ .

**Proof.** Note that  $(\text{AI}_F)$  is equivalent to the following inequalities (for all  $s, t \in T$ ):

$$U_s - U_t - \lambda_t (\mathbf{p}_t \cdot \mathbf{x}_s - m_t) \leq 0,$$

where I have defined  $m_t = \mathbf{p}_t \cdot \mathbf{x}_t - F$  for all  $t \in T$ . By Afriat's theorem, this is equivalent to that  $\mathbb{T} = \{\mathbf{p}_t, \mathbf{x}_t\}_{t \in T}$  satisfies  $\text{GARP}$ , i.e., there exists a sequence of observations  $(t, u, v, \dots, w, s) \in T$  such that:

$$\begin{aligned} \mathbf{p}_t \cdot \mathbf{x}_t - F = m_t &\geq \mathbf{p}_t \cdot \mathbf{x}_u, \quad \mathbf{p}_u \cdot \mathbf{x}_u - F = m_u \geq \mathbf{p}_u \cdot \mathbf{x}_v, \dots, \quad \mathbf{p}_w \cdot \mathbf{x}_w - F = m_w \geq \mathbf{p}_w \cdot \mathbf{x}_s \text{ implies} \\ \mathbf{p}_s \cdot \mathbf{x}_s - F = m_s &\leq \mathbf{p}_s \cdot \mathbf{x}_t. \end{aligned}$$

But this is  $\text{GARP}_F$  which proves the theorem. ■

Theorem 2 suggests that we can replace the quadratic (non-linear) constraints in the problem  $(\text{op\_AI}_F)$  with the equivalent condition  $\text{GARP}_F$  and calculate FW's test statistic,  $F$ , by solving:

$$\begin{aligned} \min_{\{F\}} F \quad & \text{s.t.} \quad (\text{op\_G}_F) \\ \mathbb{T} &= \{\mathbf{p}_t, \mathbf{x}_t\}_{t \in T} \text{ satisfies } \text{GARP}_F, \\ F &\geq 0. \end{aligned}$$

This problem can be efficiently (i.e., in polynomial time) solved using the following simple binary search algorithm.

#### Algorithm 1

*Input:* Data  $\mathbb{T} = \{\mathbf{p}_t, \mathbf{x}_t\}_{t \in T}$ , a lower bound  $F_l$ , an upper bound,  $F_u$ , and a termination criterion  $\psi > 0$ .

*Output:* A slack term  $F$  which satisfies  $\text{GARP}_F$ .

1. If  $\mathbb{T}$  satisfies  $\text{GARP}$ , abort and return  $F = 0$ . Otherwise set  $F_u^{(1)} = F_u, F_l^{(1)} = F_l$  and  $F^{(0)} = (F_u - F_l) / 2$ .

2. Set  $i = 1$ : do until  $\|F^{(i)} - F^{(i-1)}\| \leq \psi$ ,

- if  $\mathbb{T}$  satisfies  $\text{GARP}_F$  with  $F^{(i)} = (F_u^{(i)} - F_l^{(i)})/2$ , set  $F_u^{(i+1)} = (F_u^{(i)} - F_l^{(i)})/2$  and  $F_l^{(i+1)} = F_l^{(i)}$ ;
- otherwise set  $F_l^{(i+1)} = (F_u^{(i)} - F_l^{(i)})/2$  and  $F_u^{(i+1)} = F_u^{(i)}$ .

To implement this algorithm, there are a few issues to decide upon. First, we need to set lower and upper bounds. From the restriction  $F \geq 0$ , we have that the lower bound must satisfy  $F_l = 0$ . A feasible upper bound is found by noting that  $\text{GARP}_F$  is trivially true whenever  $\mathbf{p}_t \cdot \mathbf{x}_t - F = m_t \leq 0$  holds for all  $t \in T$ ; thus  $F_u = \max_{t \in T} \{\mathbf{p}_t \cdot \mathbf{x}_t\}$ . Secondly, we need to choose a suitable finite dimensional distance metric  $\|\cdot\|$ , and a sufficiently small termination criterion  $\psi > 0$ . In the empirical application, I choose  $\|\cdot\|$  to be the euclidian norm and use  $\psi = 10^{-8}$ .

Finally, it is interesting to note that the problem ( $\text{op\_G}_F$ ) is closely related to calculating the Afriat Critical Cost Efficiency Index (CCEI). Specifically, Afriat (1972) and Varian (1990) argued that fully efficient utility maximizing behavior may be a too restrictive hypothesis in empirical applications of revealed preference. For this reason, they argued that 'nearly efficient' optimizing behavior may be an equally as plausible hypothesis as efficient optimizing behavior. To allow for inefficiency in the consumers choices, Varian (1990), based on Afriat (1972), suggested to introduce a parameter  $0 \leq e \leq 1$ , such that  $\mathbf{x}_t R_e^D \mathbf{x}_s$  if  $e \mathbf{p}_t \cdot \mathbf{x}_t \geq \mathbf{p}_s \cdot \mathbf{x}_s$ .  $\text{GARP}_e$  holds if  $\mathbf{x}_t R_e \mathbf{x}_s$  implies  $e \mathbf{p}_s \cdot \mathbf{x}_s \leq \mathbf{p}_s \cdot \mathbf{x}_t$ , where  $\mathbf{x}_t R_e \mathbf{x}_s$  is the transitive closure of the relation  $R_e^D$ . CCEI is then defined as the largest possible value of  $e$ , i.e.,

$$\begin{aligned} \text{CCEI} &= \max_{\{e\}} e \quad s.t. \\ \mathbb{T} &= \{\mathbf{p}_t, \mathbf{x}_t\}_{t \in T} \text{ satisfies } \text{GARP}_e, \\ 0 &\leq e \leq 1. \end{aligned}$$

In other words, CCEI is the smallest proportion of the consumer's budget which (s)he is allowed to waste through inefficient consumption behavior. Thus, like the problem ( $\text{op\_G}_F$ ), CCEI is designed to find the smallest perturbation of total expenditure such that the data satisfies revealed preference. However, while total expenditure is additively perturbed in ( $\text{op\_G}_F$ ) (i.e. as  $\mathbf{p} \cdot \mathbf{x} - F$ ), the perturbation enters proportionally in CCEI (i.e. as  $e \mathbf{p} \cdot \mathbf{x}$ ). As a final remark, it is standard practice in empirical applications of consumer rationality to report CCEI and there exist very efficient (polynomial time) algorithms to implement CCEI. With just a minor (and simple) change in the computer code, it is possible to modify it to solve ( $\text{op\_G}_F$ ) instead of CCEI.

## 4 Implementation

The calculation of the test statistic,  $F$ , constitutes the first step in FW's measurement error procedure. This section discuss the additional steps required to test the hypothesis (HYP). But before the step-wise procedure is fully operational, there are two issues that remain to be addressed. First, the structure and distribution of the errors need to be determined. As explained above, I assume, for now, a simple classical additive measurement error structure, defined by (CA). Moreover, I assume that the errors,  $\varepsilon_t$ , are independently normally distributed random variables with mean zero and constant variance  $\sigma^2$ , i.e.  $\varepsilon_{kt} \sim N(0, \sigma^2)$ .<sup>6</sup> At this point, it is worth pointing out that the procedure is amenable under any (parametric) error structure and error distribution. In this section, I choose a simple classical additive structure with normally distributed errors for illustration, but show in the next section how the

<sup>6</sup>That is, I assume  $E[\varepsilon_{kt}\varepsilon_{js}] = 0$  for all  $(j \neq k) \in K$  and  $(s \neq t) \in T$ .

procedure can accomodate a classical multiplicative measurement error structure as well as additive and multiplicative Berkson error structures with any predetermined error distribution.

Secondly, the test statistic,  $F$ , need to be linked to a statistical decision rule in order to test the hypothesis (HYP). Given the assumption of a classical additive error structure (CA), a critical value can be derived from the following theorem (See Jones and Edgerton, 2007, p.218).

**Theorem 3** *Let  $\hat{F}$  be the optimal solution from  $(op\_AI_F)$ , or equivalently, from  $(op\_G_F)$ . Suppose (CA) holds. Then under  $H_0$  in (HYP), it holds that  $\hat{F} \leq \max_{s,t \in T} \{\mathbf{p}_t \cdot (\boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_s)\}$ .*

The decision rule is to reject  $H_0$  in (HYP) whenever  $\hat{F} > C_{1-\alpha}^{CA}$ , where  $\alpha$  denotes the significance level and  $C_{1-\alpha}^{CA}$  denotes the  $1 - \alpha$  percentile of the distribution of  $\max_{s,t \in T} \{\mathbf{p}_t \cdot (\boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_s)\}$ . However,  $\max_{s,t \in T} \{\mathbf{p}_t \cdot (\boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_s)\}$  does not follow any standard distribution, and it is therefore difficult to derive an analytical expression for  $C_{1-\alpha}$ . To deal with this, FW suggested calculating the empirical distribution of  $\max_{s,t \in T} \{\mathbf{p}_t \cdot (\boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_s)\}$  by simulations. With  $(op\_AI_F)$  replaced by the more efficient Algorithm 1, FW's measurement error procedure takes the following steps (Here and below,  $\alpha$  denotes the nominal (%)-significance level set by the researcher).

#### FW's measurement error procedure with classical additive errors

1. Calculate  $F$  using Algorithm 1. Denote the solution  $\hat{F}$ .
2. Choose  $\sigma^2$  and  $M$  and set  $m = 0$ .
3. Draw random numbers  $\varepsilon_{kt} \sim N(0, \sigma^2)$  for all  $k \in K$  and  $t \in T$ .
4. If  $\hat{F} > \max_{s,t \in T} \{\mathbf{p}_t \cdot (\boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_s)\}$ , then set  $m = m + 1$ .
5. Repeat steps 3 and 4  $M$  times.
6.  $H_0$  in (HYP) (i.e., that  $\{\mathbf{p}_t, \mathbf{q}_t\}_{t \in T}$  satisfies GARP) is rejected if  $100 \times (m/M) > (100 - \alpha)$ .

FW recommended that the number of simulations, given by  $M$  in step 2, should be set relatively large. In our application, we have  $M = 5,000$  which should be sufficient to provide a good approximation of the empirical distribution of  $\max_{s,t \in T} \{\mathbf{p}_t \cdot (\boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_s)\}$ .

Finally, step 2 require the researcher to set a predetermined value of the variance  $\sigma^2$ . In practice, Jones and Edgerton (2009) suggested implementing the procedure for a grid of values of  $\sigma^2$ . The smallest value of  $\sigma^2$  such that  $H_0$  cannot be rejected is a lower bound of the possible 'amount' of measurement errors (measured by the variance). More precisely, if  $\bar{\sigma}^2$  denotes the lower bound, then the procedure is also unable to reject  $H_0$  at the given significance level for any  $\sigma^2 \geq \bar{\sigma}^2$ .

## 5 Extensions

This section illustrates the generality of FW's procedure by showing how it can accomodate other measurement error models such as a classical multiplicative error structure and Berkson additive and multiplicative error structures. By the end of the section, I stress that the procedure can take on any imaginable (parametric) error distribution, and discuss other properties.

**Classical multiplicative error structure.** Instead of the classical additive error structure (CA), I now assume that the measurement errors,  $\varepsilon_t$  (which I suppose have mean zero), enter multiplicatively which results in the classical multiplicative error model:

$$\mathbf{x}_t = \mathbf{q}_t \odot (\mathbf{1} + \varepsilon_t), \quad (\text{CM})$$

where,  $\mathbf{1} = (1, \dots, 1)$  denotes a  $K$ -dimensional vector of ones, and  $\odot$  denotes the Hadamard product (element-wise product). Further recall that  $\{\mathbf{q}_t\}_{t \in T}$ , denotes the 'true' (unobserved and without errors) quantity-vector. In order to test the hypothesis (HYP) given the model (CM), one needs to link the test statistic,  $F$ , to a statistical decision rule. To do so, I prove the following result, which is analogous to Theorem 3 in the case of the classical multiplicative error model. Here,  $\div$  denotes Hadamard division (element-wise division).

**Theorem 4** *Let  $\hat{F}$  be the optimal solution from  $(op\_G_F)$ , or equivalently, from  $(op\_AI_F)$ . Suppose (CM) holds. Then under  $H_0$  in (HYP), it holds that*

$$\hat{F} \leq \max_{s, t \in T} \{\mathbf{p}_t \cdot ([\mathbf{x}_s \div (\mathbf{1} + \varepsilon_s)] - [\mathbf{x}_t \div (\mathbf{1} + \varepsilon_t)]) - \mathbf{p}_t \cdot (\mathbf{x}_s - \mathbf{x}_t)\}.$$

**Proof.** Define  $B = \max_{s, t \in T} \{\mathbf{p}_t \cdot ([\mathbf{x}_s \div (\mathbf{1} + \varepsilon_s)] - [\mathbf{x}_t \div (\mathbf{1} + \varepsilon_t)]) - \mathbf{p}_t \cdot (\mathbf{x}_s - \mathbf{x}_t)\}$ . By Afriat's theorem (Theorem 1) and under  $H_0$  in (HYP) there exist numbers  $U_t$  and  $\lambda_t > 0$  satisfying the (Afriat) inequalities (for all  $s, t \in T$ ):  $U_s - U_t \leq \lambda_t \mathbf{p}_t \cdot (\mathbf{q}_s - \mathbf{q}_t)$ . Now, first solving for  $\mathbf{q}_t$  in (CM) to get  $\mathbf{q}_t = \mathbf{x}_t \div (\mathbf{1} + \varepsilon_t)$ , and then substituting this into the Afriat inequalities yields  $U_s - U_t \leq \lambda_t \mathbf{p}_t \cdot (\mathbf{x}_s \div (\mathbf{1} + \varepsilon_s) - \mathbf{x}_t \div (\mathbf{1} + \varepsilon_t))$ . Next, dividing through by  $\lambda_t$  and subtracting  $\mathbf{p}_t \cdot (\mathbf{x}_s - \mathbf{x}_t)$  from both sides gives:

$$\frac{U_s - U_t}{\lambda_t} - \mathbf{p}_t \cdot (\mathbf{x}_s - \mathbf{x}_t) \leq \mathbf{p}_t \cdot (\mathbf{x}_s \div (\mathbf{1} + \varepsilon_s) - \mathbf{x}_t \div (\mathbf{1} + \varepsilon_t)) - \mathbf{p}_t \cdot (\mathbf{x}_s - \mathbf{x}_t) \leq B.$$

It, therefore, follows that (for all  $s, t \in T$ ):

$$U_s - U_t - \lambda_t \mathbf{p}_t \cdot (\mathbf{x}_s - \mathbf{x}_t) \leq \lambda_t B.$$

Since  $\hat{F}$  is the value of  $F$  that solves the problem  $(op\_AI_F)$ , or equivalently, solves  $(op\_G_F)$ , it holds that  $\hat{F} \leq B$ . ■

Similar to the case of a classical additive error structure, the statistical decision rule here is to reject  $H_0$  in (HYP) whenever  $\hat{F} > C_{1-\alpha}^{CM}$  where  $\alpha$  denotes the significance level, and  $C_{1-\alpha}^{CM}$  denotes the  $1 - \alpha$  percentile of the distribution of  $\max_{s, t \in T} \{\mathbf{p}_t \cdot ([\mathbf{x}_s \div (\mathbf{1} + \varepsilon_s)] - [\mathbf{x}_t \div (\mathbf{1} + \varepsilon_t)]) - \mathbf{p}_t \cdot (\mathbf{x}_s - \mathbf{x}_t)\}$ . Once again, we have to resort to simulations in order to calculate the empirical distribution of  $C_{1-\alpha}^{CM}$ . In this case, FW's measurement error procedure takes the following steps, assuming that the errors are normally distributed.

#### FW's measurement error procedure with classical multiplicative errors

1-3. *Same as above.*

4. *If  $\hat{F} > \max_{s, t \in T} \{\mathbf{p}_t \cdot ([\mathbf{x}_s \div (\mathbf{1} + \varepsilon_s)] - [\mathbf{x}_t \div (\mathbf{1} + \varepsilon_t)]) - \mathbf{p}_t \cdot (\mathbf{x}_s - \mathbf{x}_t)\}$ , then set  $m = m + 1$ .*

5-6. *Same as above.*

As in the previous case, that the researcher need to set a predetermined value of the variance  $\sigma^2$ . Like before, I suggest following Jones and Edgerton (2009) and implement the procedure for a grid of values of  $\sigma^2$ .

Finally, it is interesting to note that it is only step 4 in the procedure that is different from the implementation under classical additive errors. As such, neither implementation is more computationally burdensome than the other, and can easily be combined in empirical applications.

**Berkson additive error structure.** In the next two extensions, I show how to allow for Berkson error structures in FW's measurement error procedure. I begin with a Berkson additive error structure, and then turn to the multiplicative case. Suppose that the 'true' quantities  $\mathbf{q}_t$  are predicted (or caused) by the observed quantities  $\mathbf{x}_t$  via the following Berkson specification:

$$\mathbf{q}_t = \mathbf{x}_t + \boldsymbol{\eta}_t, \quad (\text{BA})$$

where  $\boldsymbol{\eta}_t$  in this case denotes the measurement errors. In contrast to classical measurement errors, a Berkson structure thus assumes that  $\mathbf{q}_t$  fluctuates around  $\mathbf{x}_t$  such that  $E[\mathbf{q}_t] = \mathbf{x}_t$ , which means that the observed quantities predict the 'true' quantities. Given the Berkson additive error structure (BA), and by defining  $\boldsymbol{\varepsilon}_t = -\boldsymbol{\eta}_t$  (so that  $\mathbf{x}_t = \mathbf{q}_t + \boldsymbol{\varepsilon}_t$ ), we can directly apply Theorem 3 and show that it under  $H_0$  in (HYP) holds that  $\hat{F} \leq \max_{s,t \in T} \{\mathbf{p}_t \cdot (\boldsymbol{\eta}_s - \boldsymbol{\eta}_t)\}$ . Thus, the decision rule in this case is to reject  $H_0$  in (HYP) whenever  $\hat{F} > C_{1-\alpha}^{BA}$  where  $C_{1-\alpha}^{BA}$  denotes the  $1 - \alpha$  percentile of the distribution of  $\max_{s,t \in T} \{\mathbf{p}_t \cdot (\boldsymbol{\eta}_s - \boldsymbol{\eta}_t)\}$ . Assuming normally distributed errors with variance  $\sigma^2$ , FW's measurement error procedure takes the following steps.

#### FW's measurement error procedure with Berkson additive errors

1-2. *Same as above.*

3. *Draw random numbers  $\eta_{kt} \sim N(0, \sigma^2)$  for all  $k \in K$  and  $t \in T$ .*

4. *If  $\hat{F} > \max_{s,t \in T} \{\mathbf{p}_t \cdot (\boldsymbol{\eta}_s - \boldsymbol{\eta}_t)\}$ , then set  $m = m + 1$ .*

5-6. *Same as above.*

**Berkson multiplicative error structure.** I finally consider a Berkson multiplicative error structure,

$$\mathbf{q}_t = \mathbf{x}_t \odot (\mathbf{1} + \boldsymbol{\eta}_t). \quad (\text{BM})$$

This is the error structure considered by Varian (1985) and is the most commonly used model in empirical applications (See the citations in the Introduction). Assuming that (BM) holds and defining  $\boldsymbol{\varepsilon}_t = -\mathbf{x}_t \odot \boldsymbol{\eta}_t$ , we can once more apply Theorem 3 and show that it holds under  $H_0$  in (HYP) that  $\hat{F} \leq \max_{s,t \in T} \{\mathbf{p}_t \cdot (\mathbf{x}_s \odot \boldsymbol{\eta}_s - \mathbf{x}_t \odot \boldsymbol{\eta}_t)\}$ . Jones and Edgerton (2009, p. 225-227) provides a detailed discussion on how to implement FW's measurement procedure given the Berkson multiplicative error model (BM) with normally distributed errors. However, note that they suggest to calculate the test statistic,  $\hat{F}$ , by solving (op\_AIF) using non-linear optimization techniques, whereas I recommend calculating it using the much faster Algorithm 1.

**Error distribution and errors in prices.** An important property of the measurement errors procedures described in the previous and this section are that they can be implemented with any imaginable (parametric) error distribution. More precisely, although we have assumed normal errors (specified in step 3), they can, in fact, take on any distribution chosen by the researcher. For example, FW assumed that the errors were uniformly distributed in their original procedure, while Cherchye, Demuynck, De Rock and Hjertstrand (2012) showed how to implement FW's measurement error procedure with normal and uniform errors in the context of a weakly separable utility function. In the empirical application below I implement FW's procedure with both uniformly and normally distributed errors to investigate the robustness of the procedures with respect to different error distributions. Interestingly, I find that the procedures seem rather robust to the choice of error distribution.

Finally, I stress that the procedures described above are not pertained to errors in the quantities. Jones and Edgerton (2009, p. 219-220), for example, describe how FW's procedure can be implemented

assuming a classical additive error specification in the prices, i.e.,  $\mathbf{p}_t = \tilde{\mathbf{p}}_t + \boldsymbol{\varepsilon}_t$ , where  $\tilde{\mathbf{p}}_t$  denotes the 'true' unobserved price-data. It is straightforward to modify FW's measurement error procedure to account for errors in the prices under any of the error models presented above.

## 6 Empirical application

This section illustrates FW's measurement error procedure under different measurement error models and distributions. I do so by applying them to a data set that has previously been used in Jones and de Peretti (2005) to compare and contrast Varian's (1985) error procedure with an alternative procedure proposed by de Peretti (2005). Both of these procedures are based on computing perturbed quantity data that satisfy **GARP**. Jones and Edgerton (2009) used the same data set to compare FW's measurement error procedure with Varian's and de Peretti's procedures. However, they exclusively considered a Berkson multiplicative error model, whereas I illustrate FW's procedure under both classical and Berkson error structures. In addition, I calculate the test statistic  $F$  using Algorithm 1, while Jones and Edgerton (2009) used non-linear optimization techniques to calculate  $F$  from the problem (op\_AI<sub>F</sub>).

The data consists of nominal per-capita asset stocks and real user cost prices for the assets in the monetary aggregate L (Liquid Assets) (See Jones and de Peretti, 2005). The data span monthly observations from 1960 to 1992, but because of inconsistencies in the data, Jones and de Peretti (2005) split the data into 8 different (non-overlapping) sub sets, which they called S1-S8. Table 1 presents the results and some summary statistics of S1-S8<sup>7</sup>. In this table, columns 4 and 5 report the number of **GARP** violations and the Afriat critical cost efficiency index (CCEI). Column 6 report the calculated test statistic,  $\hat{F}$ , for each sub-sample, while the last column gives the time (in seconds) it took to calculate  $\hat{F}$  using Algorithm 1.

Table 1: Summary statistics and results

Sample	#Assets ( $K$ )	#Obs. ( $T$ )	# GARP violations	CCEI	$\hat{F}$	Time (in sec.)
S1	13	36	2	0.9999	0.0047	1.0140
S2	14	81	1479	0.9740	4.3145	4.8828
S3	17	45	6	0.9988	0.2828	1.3104
S4	19	34	8	0.9990	0.3636	0.7020
S5	20	70	442	0.9825	14.9133	3.2604
S6	24	39	6	0.9970	0.1767	0.9672
S7	22	65	18	0.9984	0.9240	3.1512
S8	20	16	0	1.0000	0.0000	0.0000

The last column shows that Algorithm 1 indeed runs very fast in practice. Take for example the sample S2 with  $T = 81$  observations, for which Algorithm 1 found a solution under 5 seconds.

Consider next Figures 1-7, which plots the percentage number of times  $\hat{F}$  exceed the critical value ( $y$ -axis) for different values of the standard deviation ( $x$ -axis).

[FIGURES 1-7 HERE]

<sup>7</sup>See also Tables 1,2 and 3 in Jones and de Peretti (2005).

Each figure represents a different sample (S1-S7), and contains 4 different plots which corresponds to the classical multiplicative model (CM) with either normal or uniform errors, and the Berkson multiplicative model (BM) with normal and uniform errors. For example, the solid line in Figure 1 corresponding to the classical multiplicative error model (CM) with normal errors gives the %-fraction for which  $\hat{F} > \max_{s,t \in T} \{\mathbf{p}_t \cdot ([\mathbf{x}_s \div (\mathbf{1} + \boldsymbol{\varepsilon}_s)] - [\mathbf{x}_t \div (\mathbf{1} + \boldsymbol{\varepsilon}_t)]) - \mathbf{p}_t \cdot (\mathbf{x}_s - \mathbf{x}_t)\}$  for different standard deviations. Assuming a 5% nominal significance level then means that  $H_0$  in (HYP) is rejected for standard deviations that give values of that fraction above  $(100 - 5) = 95\%$ . The results from the Berkson multiplicative model with normal errors replicates the ones in Jones and Edgerton (2009, Table 2).

Some interesting findings emerge from Figures 1-7. First of all, the figures indicate that  $H_0$  (i.e., that the true data satisfy **GARP**) can only be rejected for very small standard deviations. More precisely, if the standard deviation of measurement errors is 0.9% or greater, then the test is unable to reject  $H_0$  at the 5% level for any data set. This means that the data would have to be measured very precisely in order to consider rejecting the null hypothesis of utility maximization. And as seen from the figures, this holds irrespectively of the sample (i.e., data set), the model or error distribution in question. In fact, it is interesting to note from the figures that the classical and Berkson error models (with the same error distribution) produce very similar results. This implies that FW's measurement error procedure, at least for our data set, seem to be robust in terms of the measurement error model.

Secondly, the results also seem robust in terms of the chosen error distribution. This can be seen from comparing the results of the classical and Berkson models with normal errors to those with uniform errors. In fact, they produce very similar results. Finally, it is interesting to compare our results with Jones and de Peretti (2005) who applied Varian's (1985) measurement error procedure to the same data set (assuming a Berkson multiplicative error model). Looking at Table 5 in Jones and de Peretti (2005), we see that Varian's (1985) and our results are fairly consistent with each other (which confirms the findings in Jones and Edgerton, 2009, who compared FW's and Varian's procedure within a similar setting).

## 7 Discussion and concluding remarks

This paper has extended Fleissig and Whitney's (2005) measurement error procedure to make it operational with large data sets. In particular, I have proposed an easy-to-apply algorithm to calculate the test statistic in this procedure. The algorithm is based on a simple modification of the Generalized Axiom of Revealed Preference (**GARP**), and as such, doesn't require the use of any optimization software. In an empirical application, I show that implementing this algorithm is very fast in practice.

In addition to proposing the new algorithm I have shown how Fleissig and Whitney's (2005) measurement error procedure can be applied to different measurement error models. Specifically, I show how the procedure can be designed to test whether data with errors satisfy rational behavior in classical additive and multiplicative error models as well as in Berkson additive and multiplicative error models. The measurement errors are not pertained to a particular distributional assumption in any of these models, but can take on any imaginable distribution chosen by the researcher.

Finally, I discuss some important issues related to the nature of FW's test procedure. First, it is not hard to see that the procedures described in previous sections are conservative. This follows because the test will have at least the desired level of significance, i.e., the probability of a Type I error is  $P[\hat{F} > C_{1-\alpha}^l \mid H_0] \leq \alpha$  for  $l = CA, CM, BA, BM$ , where  $\alpha$  is the nominal significance level. This means that one should expect the probability of making a Type I error to be low in the current context. But although this is not negative per sé, it is likely that it will have effect on the power of the test. Unfortunately, measuring the potential loss in power is difficult since it requires the researcher to specify a suitable hypothesis of irrational behavior. Nevertheless, the conservative nature and its implications

on the power is not unique to FW's procedure but shared by the other procedures designed to test the hypothesis (HYP); See for examples Varian (1985) and Epstein and Yatchew (1985).

Secondly, when put in relation to the alternative procedures (notably Varian, 1985, and Epstein and Yatchew, 1985) that has been developed to test the hypothesis (HYP), it seem that FW's procedure, in combination with the results in this paper, has two main benefits. First, from a computational viewpoint, it is by far the most simple and efficient to implement and one can expect it to run very fast in practice. As such, it seem to be the only procedure that is applicable for the large data sets typically encountered in empirical macroeconomics and applied economic research. Second, FW's procedure is very flexible in that it can be implemented under a wide variety of different measurement error models without increasing or changing the computational burden of the procedure. Also, it is very general in the sense that it is amenable under any error distribution chosen by the researcher.

## References

- [1] Afriat, S.N. (1967). The construction of a utility function from expenditure data. *International Economic Review* **8**, 67-77.
- [2] Afriat, S.N. (1972). Efficiency estimates of production functions. *International Economic Review* **13**, 568-598.
- [3] Barnett, W.A. (1980). Economic monetary aggregate: an application of index number and aggregation theory. *Journal of Econometrics* **14**, 11-48.
- [4] Barnett, W.A., E.W. Diewert and A. Zellner (2009). Introduction to measurement with theory. *Macroeconomic Dynamics* **13**, 151-168.
- [5] Belongia, M.T. (1996). Measurement matters: Recent results from monetary economics reexamined. *Journal of Political Economy* **104**, 1065-1083.
- [6] Carroll, R. J., D. Ruppert, and L. A. Stefanski (1995). *Measurement error in nonlinear Models: A modern perspective*. Chapman and Hall.
- [7] Chen X., H. Hong and D. Nekipelov (2011). Nonlinear models of measurement errors. *Journal of Economic Literature* **49**, 901-937
- [8] Cherchye, L., T. Demuynck, B. De Rock and P. Hjerstrand (2012). Revealed preference tests for weak separability: an integer programming approach. Working Paper.
- [9] Crawford, I. and K. Pendakur (2013). How many types are there?. *Economic Journal* **123**, 77-95.
- [10] Diewert, W.E. (1973). Afriat and revealed preference theory. *Review of Economic Studies* **40**, 419-425.
- [11] F. Echenique, S. Lee, and M. Shum. The money pump as a measure of revealed preference violations. *Journal of Political Economy* **119**, 1201-1223.
- [12] Elger, T. and B.E. Jones (2008). Can rejections of weak separability be attributed to random measurement errors in the data? *Economics Letters* **99**, 44-47.
- [13] Epstein, L.G. and A.J. Yatchew (1985). Nonparametric hypothesis testing procedures and applications to demand analysis. *Journal of Econometrics* **30**, 149-169.

- [14] Fisher, D. and A. Fleissig (1997). Monetary aggregation and the demand for assets. *Journal of Money, Credit, and Banking* **29**, 458-475.
- [15] Fixler, D. (2009). Measurement error in national accounts. *Advances in Econometrics* **24**, *Measurement error: Consequences, Applications and Solutions*, 91-105.
- [16] Fleissig, A.R. and G.A. Whitney (2005). Testing the significance of violations of Afriat's inequalities. *Journal of Business and Economic Statistics* **23**, 355-362..
- [17] Hausman, J. Mismeasured variables in econometric analysis: problems from the right and problems from the left. *The Journal of Economic Perspectives* **15**, 57-67.
- [18] Hjertstrand, P. (2007). Food demand in Sweden: A non-parametric approach. In W. A. Barnett & A. Serletis (Eds.), *Functional Structure Inference* (pp. 157-182). International Symposia in Economic Theory and Econometrics (ISETE). London: Elsevier Science.
- [19] Jha, R. and I. Longjam (2006). Structure of financial savings during Indian economic reforms. *Empirical Economics* **31**, 861-869.
- [20] Jones, B.E., D.H. Dutkowsky and T. Elger (2005). Sweep programs and Optimal Monetary Aggregation. *Journal of Banking & Finance* **29**, 483-508.
- [21] Jones, B.E. and D.L. Edgerton (2009). Testing utility maximization with measurement errors in the data. *Advances in Econometrics* **24**, *Measurement error: Consequences, Applications and Solutions*, 199-236.
- [22] Jones, B.E. and P. de Peretti (2005). A comparison of two methods for testing the utility maximization hypothesis when quantity data is measured with error. *Macroeconomic Dynamics* **9**, 612-629.
- [23] de Peretti, P. (2005). Testing the significance of the departures from utility maximization. *Macroeconomic Dynamics* **9**, 372-397.
- [24] Swofford, J.L. and G.A. Whitney (1986). Flexible functional forms and the utility approach to the demand for money: A nonparametric analysis. *Journal of Money, Credit, and Banking* **18**, 383-389.
- [25] Swofford, J.L. and G.A. Whitney (1987). Nonparametric tests of utility maximization and weak separability for consumption, leisure and money. *Review of Economic and Statistics* **69**, 458-464.
- [26] Swofford, J.L. and G.A. Whitney (1994). A revealed preference test for weakly separable utility maximization with incomplete adjustment. *Journal of Econometrics* **60**, 235-249.
- [27] Varian, H.R. (1982). The nonparametric approach to demand analysis. *Econometrica* **50**, 945-972.
- [28] Varian, H.R. (1985). Non-parametric analysis of optimizing behavior with measurement error. *Journal of Econometrics* **30**, 445-458.
- [29] Varian, H.R. (1990). Goodness-of-fit in optimizing models. *Journal of Econometrics* **46**, 125-140.
- [30] Varian, H.R. (2006). Revealed preference. In M. Szenburg, L. Ramrattan and A.A. Gottesman (eds.) *Samuelsonian Economics and the Twenty-First Century*, 99-115. Oxford: Oxford University Press.
- [31] Warshall, S. (1962). A theorem on Boolean matrices. *Journal of the American Association for Computing Machinery* **9**, 11- 12.

Figure 1: Results for sample S1

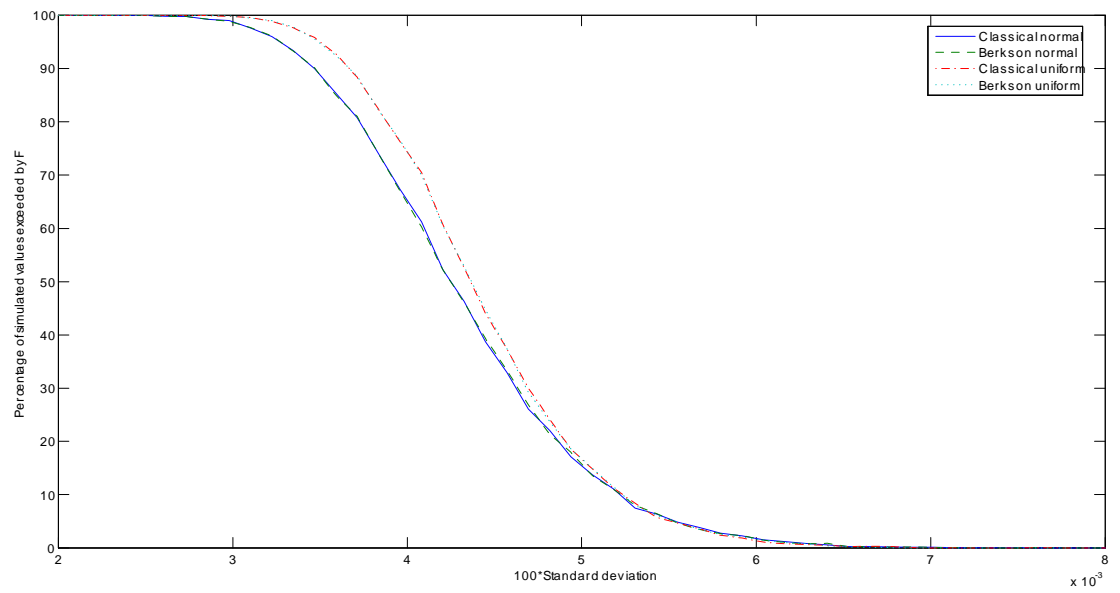


Figure 2: Results for sample S2

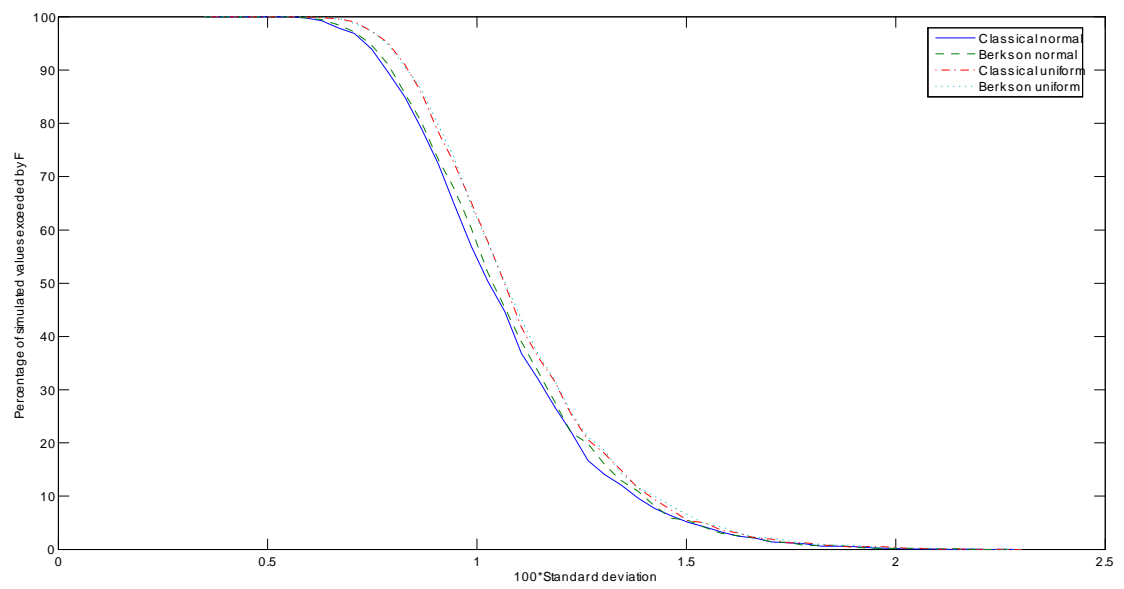


Figure 3: Results for sample S3

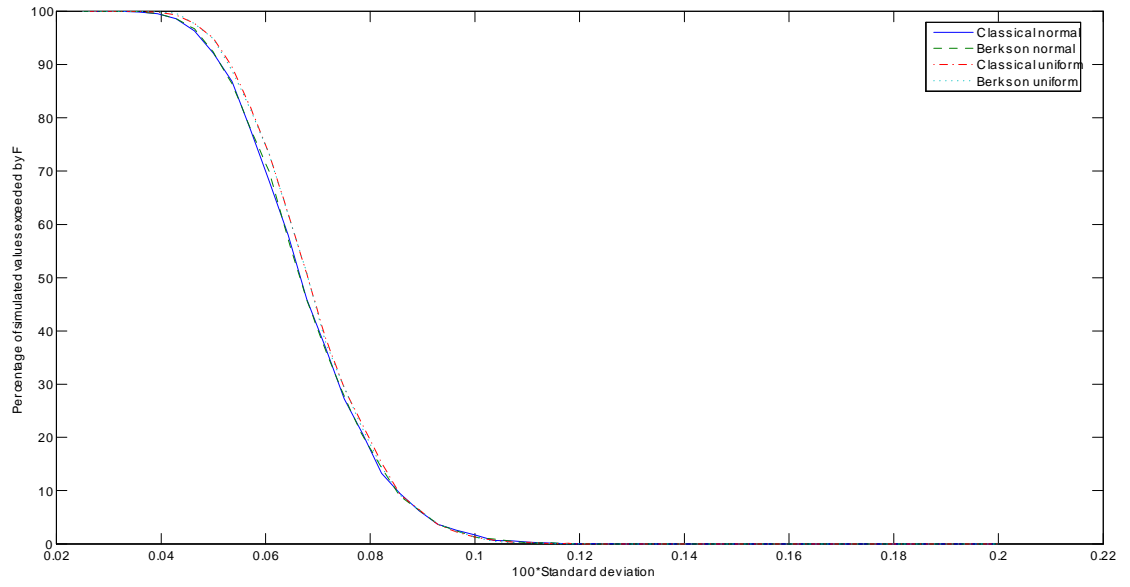


Figure 4: Results for sample S4

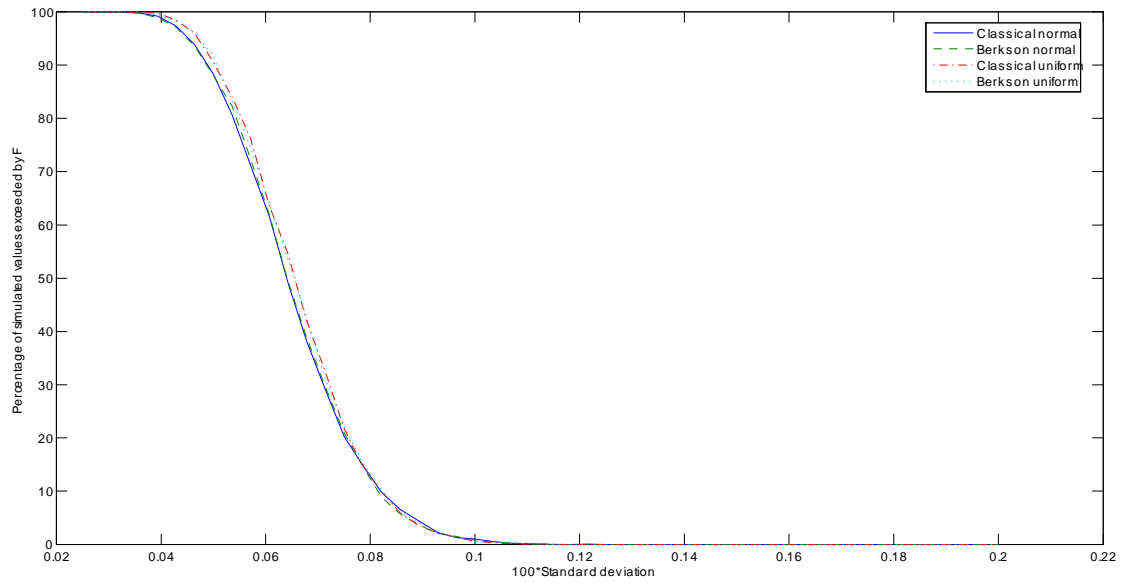


Figure 5: Results for sample S5

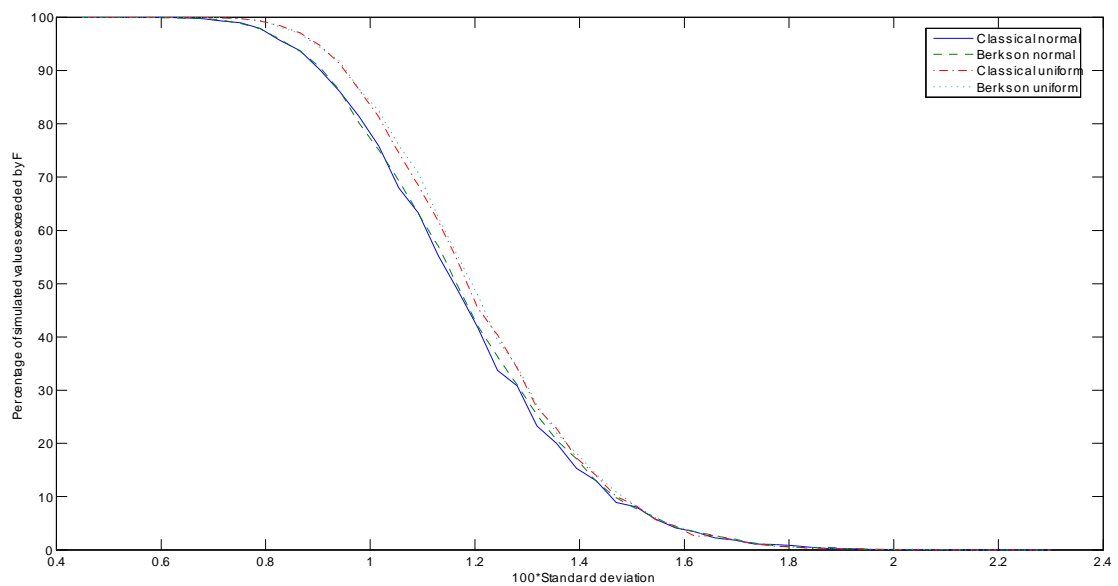


Figure 6: Results for sample S6

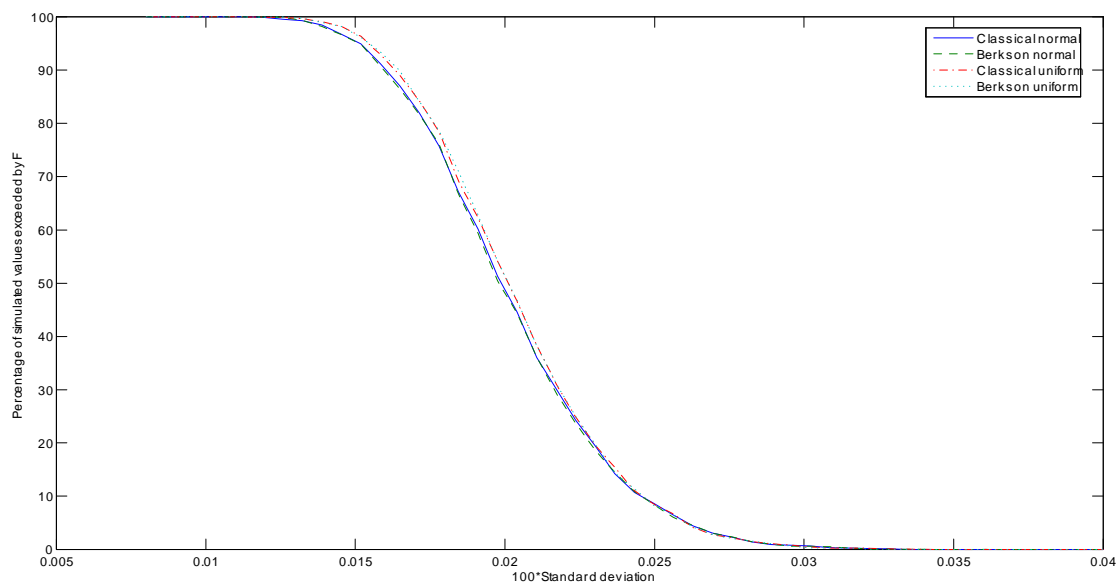


Figure 7: Results for sample S7

