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

Per Hjertstrand*

Research Institute of Industrial Economics (IFN)

Stockholm, Sweden

December 2014

Abstract

Revealed preference tests are widely used in empirical applications of consumer rationality. These tests 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 classical 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.[1]

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. Web: http://www.ifn.se/eng/people/research_fellows/per-hjertstrand

[1] 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.[2] However, these models have been labeled inappropriate to describe many economic data sets (Chen, Hong and Nekipelov, 2011).[3] 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).[4]

From an applied perspective, revealed preference based tests are attractive since they are nonparametric 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) and de Peretti (2005) are a few examples of such test-procedures. However, some of these procedures may be 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. de Peretti's (2005) procedure is based on adjusting the bundles that are directly causing the revealed preference violations. Jones and de Peretti (2005) compared Varian's (1985) and de Peretti's (2005) procedures and found the latter to be substantially faster in practice.

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

---

[2] See Varian (1985), Epstein and Yatchew (1985), Jones and de Perreti (2005), Elger and Jones (2008) and Hjertstrand (2007).

[3] 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.

[4] 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. Cherchye, Demuynck, De Rock and Hjertstrand (2014) recently proposed a new revealed preference test for weak separability based on solving a mixed integer linear programming problem.

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.

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). 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).

I show that Fleissig and Whitney's (2005) test statistic closely resembles the Afriat efficiency index (`AEI`)[5] both in terms of interpretation and computation. Specifically, I show that both are measures of the smallest perturbation of total expenditure such that observed data satisfies revealed preference. They simply differ in how expenditure is perturbed: While the `AEI` finds the smallest *proportional* perturbation, the test statistic finds the smallest *additive* perturbation. Moreover, I show that calculating the test statistic and the `AEI` is equivalent in terms of computational complexity. More specifically, both can be calculated using essentialy the same algorithm; the paper gives a detailed account of this algorithm. Reporting the AEI has become standard in empirical studies of consumer rationality. The new algorithm implementing Fleissig and Whitney's (2005) error procedure provides researchers with a computationally equivalent method to test whether their data satisfies revealed preference, given the presence of measurement errors.

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 approporiate ones in empirical economic modelling. In fact, this highlights one of the main advantages of Fleissig and Whitney's (2005) procedure: it is very flexible in the sense that it can be implemented under a wide variety of different measurement error models without increasing or changing the computational burden of the procedure. 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 normal, log-normal 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 contains the empirical application. Section 6 gives some concluding remarks. An online appendix containing applications to Berkson measurement error models and a Monte Carlo simulation study which illustrates the small-sample properties of the test procedure is downloadable from the author's webside.[6]

---

[5] The Afriat efficiency index (`AEI`) is also called the Critical cost efficiency index (`CCEI`).

[6] `http://www.ifn.se/eng/people/research_fellows/per-hjertstrand`.

2 Revealed preference

Suppose a consumer chooses from K goods observed at N time periods, with index set $T = \{1, ..., N\}$. Let $\mathbf{x}_t = (x_{1t}, ..., x_{Kt}) \in \mathbb{R}^K_+$ denote the observed quantity-vector at time $t \in T$ with corresponding price-vector $\mathbf{p}_t = (p_{1t}, ..., p_{Kt}) \in \mathbb{R}^K_{++}$. 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.,

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

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* $(t, u, v, ..., w, s) \in T$ *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* $s, t \in T$*):*

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

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.[7] The second method uses linear programming (LP) techniques to check whether there exists a solution (in the unknowns U_t and λ_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).

3 Allowing for measurement errors

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

$$\mathbf{x}_t = \mathbf{q}_t + \boldsymbol{\varepsilon}_t, \tag{CA}$$

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. The supplementary material accompying the paper considers additive and multiplicative Berkson measurement error models.[8]

Our purpose is to test the following hypothesis:

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

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 evaluated by comparing the test statistic to a critical value obtained from the empirical distribution of the errors.

[7] See Varian (2006) for a recent overview.

[8] The supplementary material is downloadable from the author's webside.

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

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

for all $s,t\in T$. The problem (op_AI$_\text{F}$) contains quadratic (non-linear) constraints, which makes it non-trival. As such, one alternative is to solve (op_AI$_\text{F}$) using optimization software that is able to handle quadratic constraints. However, a computationally simpler solution can be obtained by noting that if (op_AI$_\text{F}$) 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 below. 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, I introduce an alternative procedure to calculate this statistic.[9] 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$_\text{F}$) 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 (op_AI$_\text{F}$). For that reason, it is computationally much more efficient than solving (op_AI$_\text{F}$), 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 (GARP$_\mathbf{F}$) *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* $(t,u,v,...,w,s)\in T$ *such that* $\mathbf{x}_t R_F^D \mathbf{x}_u$, $\mathbf{x}_u R_F^D \mathbf{x}_v$, ..., $\mathbf{x}_w R_F^D \mathbf{x}_s$.
- *that* $\mathbb{T}$ *satisfies* 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 `GARP`$_\text{F}$ is equivalent to the non-linear constraints in (op_AI$_\text{F}$).

Theorem 2 *Consider the data $\mathbb{T}=\{\mathbf{p}_t,\mathbf{x}_t\}_{t\in T}$. For any 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 $s,t\in T$):*

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

[9] Fleissig and Whitney (2005) suggested another test statistic to implement (HYP). However, it is difficult to interpret their test statistic (See Jones and Edgerton, 2009, p.217, for a detailed discussion). The new algorithm can only calculate Jones and Edgerton's (2009) version of the test statistic F (corresponding to solving the non-linear problem (op_AI$_\text{F}$)).

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

Proof. *Note that* (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* $\texttt{GARP}$*, i.e., there exists a sequence of observations* $(t, u, v, ..., w, s) \in T$ *such that:*

$\mathbf{p}_t \cdot \mathbf{x}_t - F = m_t \geq \mathbf{p}_t \cdot \mathbf{x}_u,\ \mathbf{p}_u \cdot \mathbf{x}_u - F = m_u \geq \mathbf{p}_u \cdot \mathbf{x}_v, ...,\ \mathbf{p}_w \cdot \mathbf{x}_w - F = m_w \geq \mathbf{p}_w \cdot \mathbf{x}_s$ *implies* $\mathbf{p}_s \cdot \mathbf{x}_s - F = m_s \leq \mathbf{p}_s \cdot \mathbf{x}_t$.

But this is $\texttt{GARP}_\text{F}$ *which proves the theorem.* ■

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

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

This problem gives exactly the same solution as calculating F from (op_AI_F) and importantly, it can be very efficiently solved (i.e., in polynomial time) 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* $\texttt{GARP}_\text{F}$*.*

1. *If* $\mathbb{T}$ *satisfies* $\texttt{GARP}$*, 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* $\left\| F^{(i)} - F^{(i-1)} \right\| \leq \psi$,
   - *if* $\mathbb{T}$ *satisfies* $\texttt{GARP}_\text{F}$ *with* $F^{(i)} = \left(F_u^{(i)} - F_l^{(i)}\right)/2$*, set* $F_u^{(i+1)} = \left(F_u^{(i)} - F_l^{(i)}\right)/2$ *and* $F_l^{(i+1)} = F_l^{(i)}$;
   - *otherwise set* $F_l^{(i+1)} = \left(F_u^{(i)} - F_l^{(i)}\right)/2$ *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 $\texttt{GARP}_\text{F}$ 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 $\psi = 10^{-8}$.

Interpretation of the test statistic F. A close inspection of $\texttt{GARP}_\text{F}$ shows that the test statistic F can be interpreted as a measure of perturbed expenditure. Consider the revealed preferred relation $\mathbf{x}_t R_F^D \mathbf{x}_s$ in Definition 1: We say that $\mathbf{x}_t R_F^D \mathbf{x}_s$ if $\mathbf{p}_t \cdot (\mathbf{x}_t - \mathbf{x}_s) \geq F$, or equivalently expressed: $\mathbf{x}_t R_F^D \mathbf{x}_s$ if $m_t \geq \mathbf{p}_t \cdot \mathbf{x}_s$, where $m_t = \mathbf{p}_t \cdot \mathbf{x}_t - F$. If we define $m_t = \mathbf{p}_t \cdot \mathbf{x}_t - F$ for all $t \in T$, then the data $\mathbb{T} = \{\mathbf{p}_t, \mathbf{x}_t, m_t\}_{t \in T}$ satisfies $\texttt{GARP}_\text{F}$ if $\mathbf{x}_t R_F \mathbf{x}_s$ implies $m_s \leq \mathbf{p}_s \cdot \mathbf{x}_t$. Calculating the minimal F satisfying $\texttt{GARP}_\text{F}$ can therefore be interpreted as finding 'new' (maximal) expenditure levels $\{m_t\}_{t \in T}$ such that $\mathbb{T} = \{\mathbf{p}_t, \mathbf{x}_t, m_t\}_{t \in T}$ satisfies $\texttt{GARP}_\text{F}$. Thus, F can be interpreted as the minimal amount of expenditure which the consumer is allowed to 'waste' due to measurement errors in the data. Put differently, it can be viewed as the minimal cost that the consumer 'pays' for the measurement errors.

Connection to the Afriat efficiency index. FW's test statistic F closely resembles the Afriat efficiency index (AEI) both in terms of interpretation and computation. Indeed, both F and the AEI are measures of the smallest perturbation of total expenditure such that the observed data $\mathbb{T} = \{\mathbf{p}_t, \mathbf{x}_t\}_{t \in T}$ satisfies GARP - their subtle difference is in *how* total expenditure is perturbed. To better understand this, it is constructive to briefly review the AEI.

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 plausable hypothesis as efficient optimizing behavior. To allow for inefficiency in the consumer's choices, Varian (1990), based on Afriat (1972), suggested to introduce the 'inefficiency parameter' e by relaxing the revealed preference chain such that $\mathbf{x}_t R_e^D \mathbf{x}_s$ if $e\mathbf{p}_t \cdot \mathbf{x}_t \geq \mathbf{p}_t \cdot \mathbf{x}_s$. 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. The AEI is then defined as the value of e closest to 1, i.e.,

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

In words, the AEI is the smallest proportion of the consumer's budget which (s)he is allowed to waste through inefficient consumption behavior. Thus, like FW's test statistic F, the AEI finds the smallest perturbation of total expenditure such that the data satisfies revealed preference. However, while the test statistic F finds the smallest *additive* perturbation as $m = \mathbf{p} \cdot \mathbf{x} - F$, the AEI finds the smallest *proportional* perturbation as $m = e\mathbf{p} \cdot \mathbf{x}$.[10]

The close resemblance between F and AEI also carries over in terms of computation. In fact, Algorithm 1 was originally designed to calculate the AEI (Varian, 1990), and replacing GARP_F with GARP_e gives Varian's original algorithm. Since GARP_F and GARP_e are equivalent in terms of computational complexity, calculating the minimal F in the problem (op_G_F) is achieved in polynomial time. In practice, a researcher having access to computer code which calculates the AEI for a given data set can easily modify his code to instead calculate F by replacing GARP_e with GARP_F. As a final remark, it is standard practice in empirical applications of consumer rationality to report the AEI. The new algorithm implementing FW's measurement error procedure gives researchers a computationally equivalent method to also test for rationality given errors in their data.

4 Implementing the procedure

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 above, I assume a simple classical additive measurement error structure, defined by (CA). Moreover, I assume that the errors, $\boldsymbol{\varepsilon}_t$, are independently normally distributed random variables with mean zero and constant variance σ^2, i.e. $\varepsilon_{kt} \sim N(0, \sigma^2)$.[11] At this point, it is worth to point out that the procedure is amenable under any (parametric) error distribution (which is possible to simulate). Later in this section, I show how the

[10] As a referee of this paper pointed out, adjusting expenditure using the AEI in revealed preference tests may lead to low test power. The same may apply for FW's measurement error procedure. To analyze this, the supplementary material contain a small simulation study. The results indicates that the procedure has good power against unifomly random behavor even when the 'true' data are shocked with up to 7-8% measurement errors.

[11] That is, I assume $E[\varepsilon_{kt}\varepsilon_{js}] = 0$ for all $(j \neq k) = 1, ..., K$ and $(s \neq t) \in T$.

procedure can be implemented for classical multiplicative measurement error models. The supplementary material shows how to implement the procedure for additive and multiplicative Berkson error models (this material can be downloaded from the author's webside).

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 $\widehat{F}$ 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 $\widehat{F} \leq \max_{s,t \in T} \{\mathbf{p}_t \cdot (\boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_s)\}$.*

The decision rule is to reject H_0 in (HYP) whenever $\widehat{F} > C_{1-\alpha}^{CA}$, where α 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$_\mathrm{F}$) replaced by the more efficient Algorithm 1, FW's measurement error procedure takes the following steps (throughout, α 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 $\widehat{F}$.*
2. *Choose σ^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 $\widehat{F} > \max_{s,t \in T} \{\mathbf{p}_t \cdot (\boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_s)\}$, 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 $\max_{s,t \in T} \{\mathbf{p}_t \cdot (\boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_s)\}$.

Step 2 requires the researcher to set a predetermined value of the variance σ^2. In practice, Jones and Edgerton (2009) suggested implementing the procedure for a grid of values of σ^2. The smallest value of σ^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 $\overline{\sigma}^2$ denotes the lower bound, then the procedure is also unable to reject H_0 at the given significance level for any $\sigma^2 \geq \overline{\sigma}^2$. Thus, the lower bound $\overline{\sigma}^2$ is a measure of what the unknown variance of the measurement errors would have to be in order to reject utility maximization. This approach closely follows Varian (1985) who states that (on p.450): "If $\overline{\sigma}^2$ is much smaller than our prior opinions concerning the precision with which these data have been measured, we may well want to accept the maximization hypothesis". Elger and Jones (2008) advocates this approach and argues (on p.46) that the lower bound $\overline{\sigma}^2$ is "...a transparent way of reporting empirical results, since it can be easily compared to one's own subjective prior regarding the [variance] of measurement errors in the data".

Classical multiplicative error model. Instead of the classical additive error model, I assume that the measurement errors, $\boldsymbol{\varepsilon}_t$ (which I suppose have mean zero), enter multiplicatively which results in the classical multiplicative error model:

$$\mathbf{x}_t = \mathbf{q}_t \odot (\mathbf{1} + \boldsymbol{\varepsilon}_t), \tag{CM}$$

where, $\mathbf{1} = (1, ..., 1)$ 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 multiplicative 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 element-wise division.[12]

Theorem 4 *Let $\widehat{F}$ 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*

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

Proof. *Define $B = \max_{s,t\in T} \{\mathbf{p}_t \cdot ([\mathbf{x}_s \div (\mathbf{1} + \boldsymbol{\varepsilon}_s)] - [\mathbf{x}_t \div (\mathbf{1} + \boldsymbol{\varepsilon}_t)]) - \mathbf{p}_t \cdot (\mathbf{x}_s - \mathbf{x}_t)\}$. 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 $\mathbf{q}_t = \mathbf{x}_t \div (\mathbf{1} + \boldsymbol{\varepsilon}_t)$, and then substituting this into the Afriat inequalities yields $U_s - U_t \leq \lambda_t \mathbf{p}_t \cdot (\mathbf{x}_s \div (\mathbf{1} + \boldsymbol{\varepsilon}_s) - \mathbf{x}_t \div (\mathbf{1} + \boldsymbol{\varepsilon}_t))$. Next, dividing through by λ_t and subtracting $\mathbf{p}_t \cdot (\mathbf{x}_s - \mathbf{x}_t)$ from both sides gives:*

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

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 $\widehat{F}$ is the value of F that solves the problem (op_ AI_{F}), or equivalently, solves (op_ G_{F}), it holds that $\widehat{F} \leq B$. ■

Similar to the case of a classical additive error structure, the statistical decision rule here is to reject H_0 in (HYP) whenever $\widehat{F} > C^{CM}_{1-\alpha}$ where α denotes the significance level, and $C^{CM}_{1-\alpha}$ denotes the $1 - \alpha$ percentile of the distribution of $\max_{s,t\in T} \{\mathbf{p}_t \cdot ([\mathbf{x}_s \div (\mathbf{1} + \boldsymbol{\varepsilon}_s)] - [\mathbf{x}_t \div (\mathbf{1} + \boldsymbol{\varepsilon}_t)]) - \mathbf{p}_t \cdot (\mathbf{x}_s - \mathbf{x}_t)\}$. Again, we have to resort to simulations in order to calculate the empirical distribution of $C^{CM}_{1-\alpha}$. 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 $\widehat{F} > \max_{s,t\in T} \{\mathbf{p}_t \cdot ([\mathbf{x}_s \div (\mathbf{1} + \boldsymbol{\varepsilon}_s)] - [\mathbf{x}_t \div (\mathbf{1} + \boldsymbol{\varepsilon}_t)]) - \mathbf{p}_t \cdot (\mathbf{x}_s - \mathbf{x}_t)\}$, then set $m = m + 1$.*

5-6. *Same as above.*

[12] Thus, for two $K-$dimensional vectors $\mathbf{z}^1 = (z^1_1, z^1_2, ..., z^1_K)$ and $\mathbf{z}^2 = (z^2_1, z^2_2, ..., z^2_K)$, we have $\mathbf{z}^1 \div \mathbf{z}^2 = (z^1_1/z^2_1, z^1_2/z^2_2, ..., z^1_K/z^2_K)$.

As above, the researcher need to set a predetermined value of the variance σ^2; I suggest following Jones and Edgerton (2009) and implement the procedure for a grid of values of σ^2. It is interesting to note that it is only the simulation of the critial value in 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. This highlights the advantage of FW's measurement error procedure - it is applicable to a wide variety of models at the same computational complexity.

As a referee of this paper pointed out, the classical multiplicative error model (CM) has a natural interpretation and motivation in terms of log-additivity, in which case it is more natural to assume that the measurement errors are log-normally distributed. Let us start from the following intuitively reasonable log-additive setup (assuming $(\mathbf{x}_t, \mathbf{q}_t) \in \mathbb{R}^K_{++} \times \mathbb{R}^K_{++}$):

$$\ln(x_{kt}) = \ln(q_{kt}) + \eta_{kt},$$

for all $t \in T$ and all $k = 1, ..., K$, where $\eta_{kt} \sim N\left(0, \sigma^2_\eta\right)$. Solving for the observed quantities yields the equivalent multiplicative formulation:

$$\mathbf{x}_t = \mathbf{q}_t \odot e^{\boldsymbol{\eta}_t} = \mathbf{q}_t \odot \boldsymbol{\varepsilon}_t,$$

for all $t \in T$. Since $\boldsymbol{\eta}_t$ are normally distributed, $\boldsymbol{\varepsilon}_t = e^{\boldsymbol{\eta}_t}$ are log-normal with mean $e^{\sigma^2_\eta/2}$ and variance $\left(e^{\sigma^2_\eta} - 1\right) e^{\sigma^2_\eta}$. This interpretation hinges on the fact that the errors in the log-additive setup are normally distributed, since otherwise it is unlikely that $e^{\boldsymbol{\eta}_t}$ will follow a standard distribution. Nevertheless, imposing normality on $\boldsymbol{\eta}_t$ may in many applications be more sensible than imposing normality directly on $\boldsymbol{\varepsilon}_t$. In the empirical application, I apply the procedure under the assumption of normal, log-normal and uniformly distributed errors to analyze how robust the method is with respect to the error distribution.

5 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 satisfies `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 the procedure under both classical and Berkson multiplicative 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_F).

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[13]. In this table, columns 4 and 5 report the number of `GARP` violations and the Afriat efficiency index (`AEI`). Column 6 reports the calculated test statistic, denoted $\widehat{F}$, for each sub-sample.

[13]See also Tables 1,2 and 3 in Jones and de Peretti (2005).

Table 1: Summary statistics and results

Sample	# Assets (K)	# Obs. (N)	# GARP violations	AEI	$\widehat{F}$
S1	13	36	2	0.9999	0.0047
S2	14	81	1479	0.9740	4.3145
S3	17	45	6	0.9988	0.2828
S4	19	34	8	0.9990	0.3636
S5	20	70	442	0.9825	14.9133
S6	24	39	6	0.9970	0.1767
S7	22	65	18	0.9984	0.9240
S8	20	16	0	1.0000	0.0000

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

[FIGURES 1-7 HERE]

Each figure represents a different sample (S1-S7), and contains 6 different plots corresponding to: (i) the classical multiplicative model (CM) with normal, uniform and log-normal measurement errors, and (ii) the Berkson multiplicative model with normal, uniform and log-normal measurement errors. For example, the solid line marked with '+' in Figure 1 corresponds to the classical multiplicative error model (CM) with normal errors and gives the fraction for which $\widehat{F} > \max_{s,t\in T}\{\mathbf{p}_t\cdot([\mathbf{x}_s \div (\mathbf{1}+\boldsymbol{\varepsilon}_s)] - [\mathbf{x}_t \div (\mathbf{1}+\boldsymbol{\varepsilon}_t)]) - \mathbf{p}_t\cdot(\mathbf{x}_s - \mathbf{x}_t)$ 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. Let us first focus on figures 1,3,4,6 and 7 (samples S1, S3, S4, S6 and S7) and use a 5% nominal significance level throughout. If the standard deviation of measurement errors is 0.08% or larger then the procedure is unable to reject the null of utility maximization for any sample in this group. Thus, assuming normality implies that measurement errors of ±0.24% of the observed quantities lies within three standard deviations of the mean. Hence, the data would have to be measured very precisely in order to reject utility maximization. Of course, it is a subjective matter whether one believes that the data contain measurement errors of at most 0.24%, but the amount seems inarguably small. Looking at samples S1 and S6, the results are even more noticable; in these cases, one rejects the null if the standard deviation is lower than 0.02%, which implies that errors of ±0.06% of the observed quantities lies within three standard deviations of the mean.

Consider next the samples S2 and S5 in figures 2 and 5. These samples have substantially more GARP violations which, on first sight, could be taken as they contain more errors. If the standard deviation of measurement errors is lower than 1% then our test rejects the null of utility maximization for sample S5. Thus, assuming normality implies that measurement errors of ±3% of the observed quantities lies within three standard deviations of the mean. This is about 12 times larger than for any of the samples in Group 1. Although it shouldn't be taken as a definite rejection of utility maximization, in combination with the large amount of GARP violations, it can be taken as an indication of rejecting utility maximization. The null of utility maximization can be rejected for a slightly lower standard deviation of measurement errors in sample S2.

Our second main point is that it is interesting to note that the classical and Berkson error models produce very similar results (for any error distribution). 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. Finally, we see that the procedure produces very similar results for all three error distributions. Thus, the results seem very robust in terms of the chosen error distribution. One way to view this finding is that higher moments (e.g. skewness and kurtosis) seem to play little role for the results and the procedure.

6 Concluding remarks

This paper has extended Fleissig and Whitney's (2005) measurement error procedure to make it operational to 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 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. 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 $P\left[\widehat{F} > C^l_{1-\alpha} \mid H_0\right] \leq \alpha$ for $l = CA, CM$, where α 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. The conservative nature is not unique to FW's procedure but shared by the other procedures designed to test the hypothesis (HYP); See 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 is applicable for large data sets which are 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. Hjertstrand (2014). Revealed preference tests for weak separability: an integer programming approach. Forthcoming in *Journal of Econometrics*.

[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] Echenique, F., 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. (2001). Mismeasured variables in econometric analysis: problems from the right and problems from the left. *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). Nonparametic 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

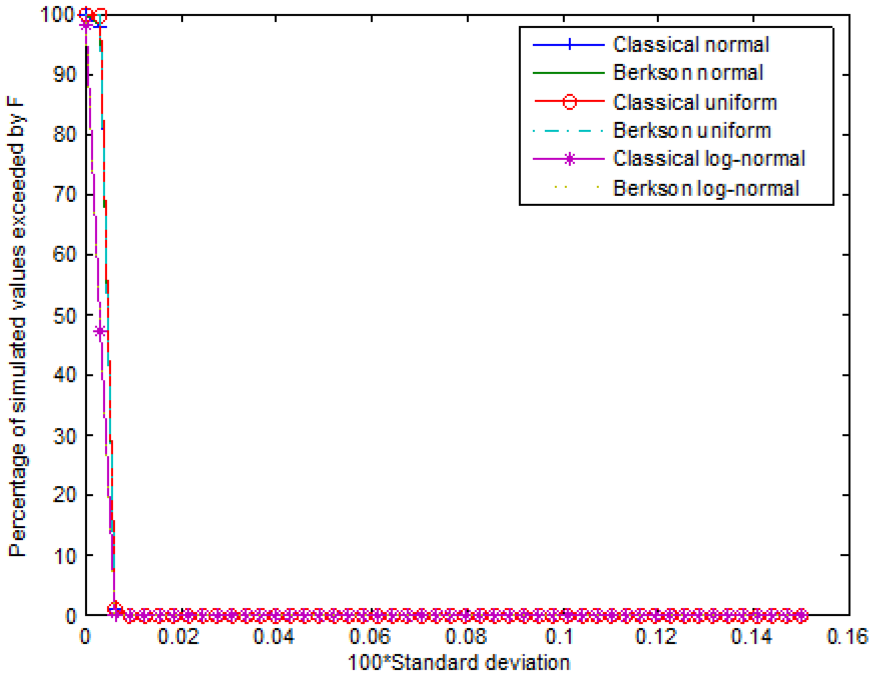

Figure 2: Results for sample S2

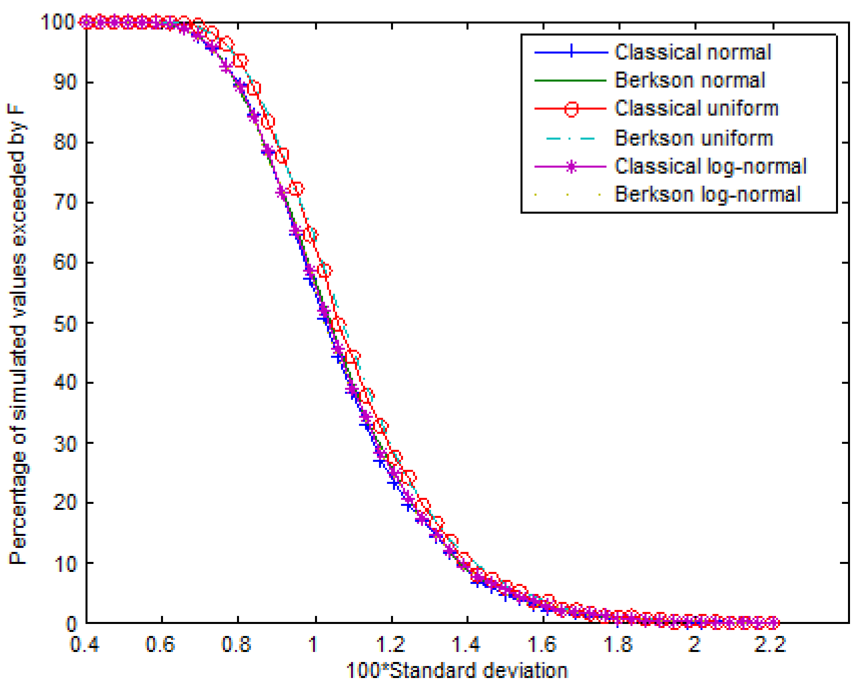

Figure 3: Results for sample S3

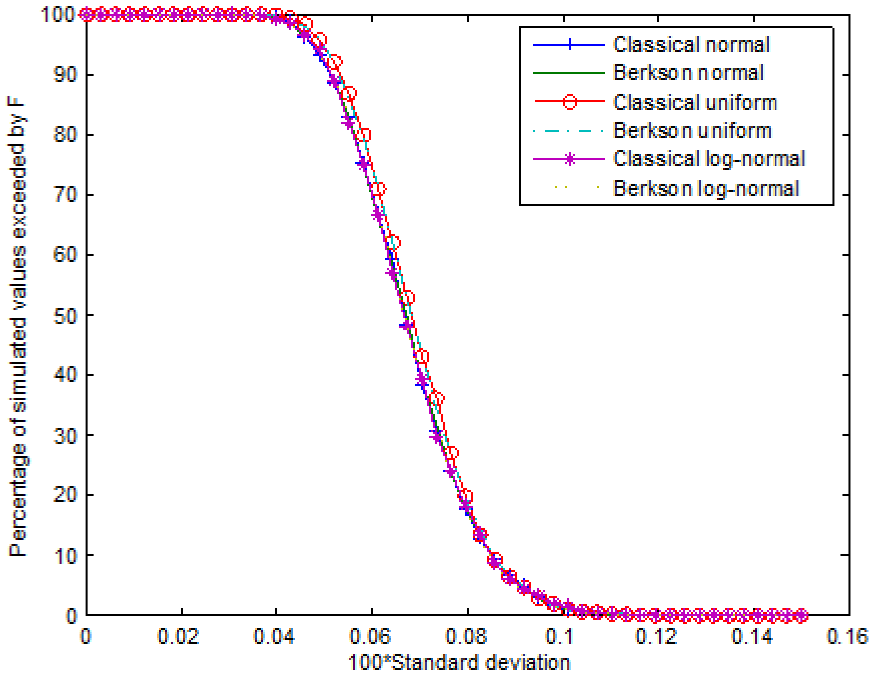

Figure 4: Results for sample S4

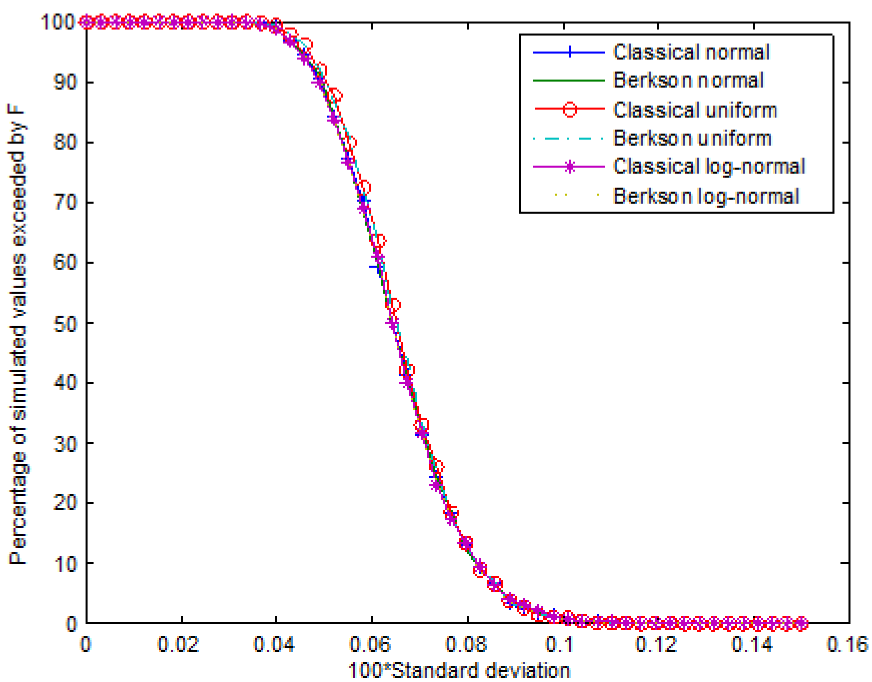

Figure 5: Results for sample S5

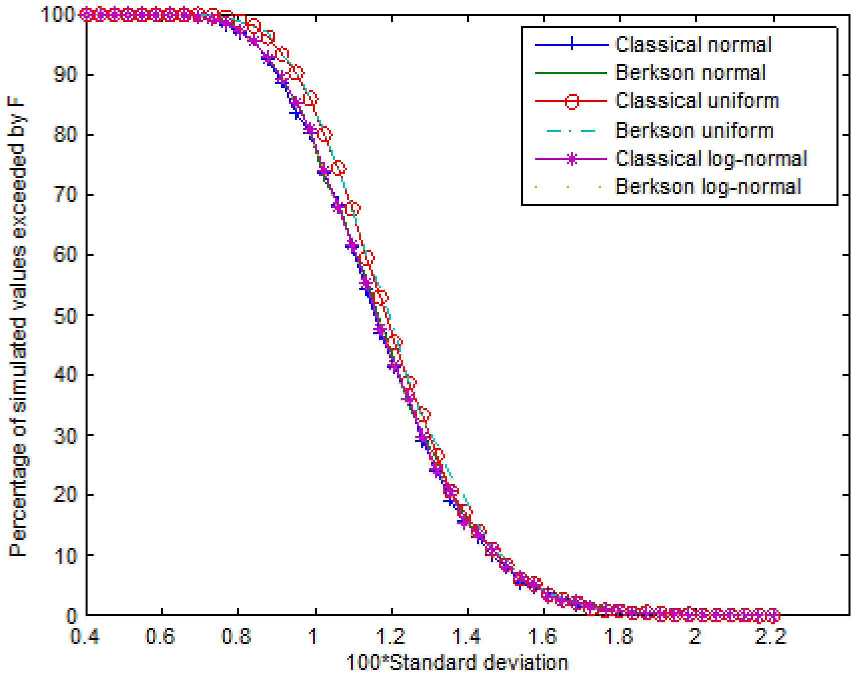

Figure 6: Results for sample S6

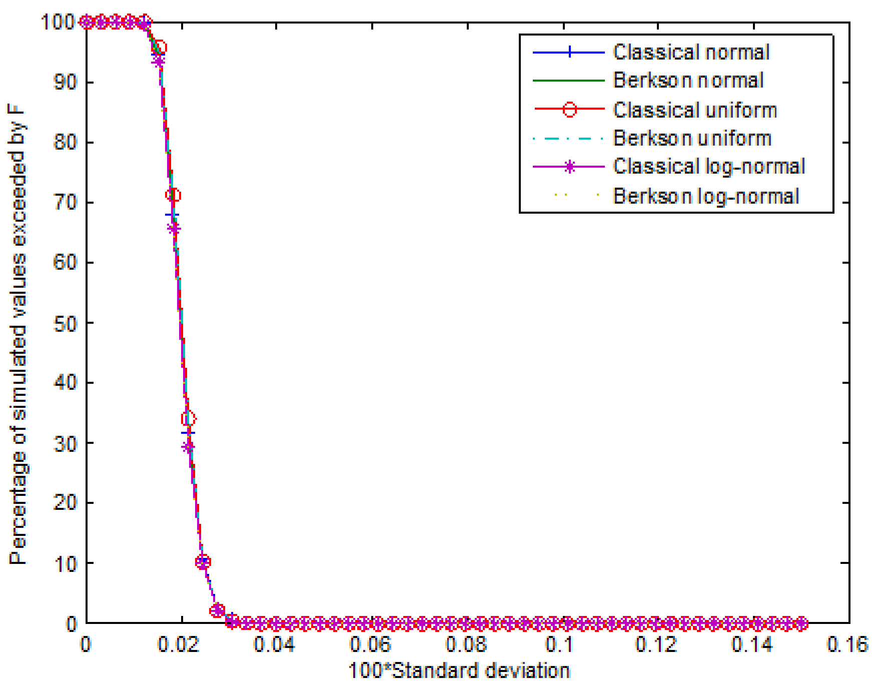

Figure 7: Results for sample S7