WORST-CASE ESTIMATION AND ASYMPTOTIC THEORY FOR MODELS WITH UNOBSERVABLES[*]

Mercedes Esteban-Bravo[1] and Jose M. Vidal-Sanz[2]

Abstract

This paper proposes a worst-case approach for estimating econometric models containing unobservable variables. Worst-case estimators are robust against the adverse effects of unobservables. In contrast to the classical literature, there are no assumptions about the statistical nature of the unobservables in a worst-case estimation. This method is robust with respect to the unknown probability distribution of the unobservables and should be seen as a complement to standard methods, as cautious modelers should compare different estimations to determine robust models. The limit theory is obtained. A Monte Carlo study of finite sample properties has been conducted. An economic application is included.

Running title: Worst-case estimation with unobservables.

Keywords: Unobservable variables, robust estimation, minimax optimization, M-estimators, GMM estimators

[1] Dept. of Business Economics, Universidad Carlos III de Madrid, C/ Madrid, 126. 28903 Getafe, Madrid, Spain. E-mail: mesteban@emp.uc3m.es

[2] Dept. of Business Economics, Universidad Carlos III de Madrid, C/ Madrid, 126. 28903 Getafe, Madrid, Spain. E-mail: jvidal@emp.uc3m.es

1 Introduction

Many theoretical models in the social sciences depend on conceptual variables that cannot be observed. For example, economic models often postulate the existence of such concepts as permanent income, expected price, human capital, personal ambition and ability. The omission of unobservables in the estimation procedures usually produces severe biases. Griliches and Mason (1972), Chamberlain and Griliches (1975) Chamberlain (1977), and Griliches (1977) have all noticed the bias inherent in income-education regressions caused by the omission of unobservable variables measuring initial ability.

Following the work of Griliches (1974) and Goldberger (1974), a large body of econometrics and statistics literature has addressed the estimation of models containing unobservables. The estimation of these models requires the assumption of a structure for the unobservables (either assuming a probability distribution or considering proxy variables and postulating a measurement error model). The consistency of these estimators is conditioned to the validity of the postulated hypotheses. Nonobservability renders the diagnosis of these hypotheses difficult to implement, even though its fulfilling will be crucial to reject an economic theory as false for contradicting empirical analysis. This limitation sometimes leads to testing theories based on facts to which they were not meant to be applied. Hence, it seems relevant to study robust estimation approaches in order to deal with unobservables.

One of the most popular techniques is the use of proxies with an "errors in variables model". Aigner (1974) uses the 1967 Survey of Economic Opportunities (SEO) to estimate the labor-supply function as an errors in variables model. Hum and Simpson (1994) suggest that a bias in labour-supply estimation is caused by the omission of such unobservable individual variables as ambition and preferences. Attempts to solve this problem using household wealth as a proxy are unsatisfactory because wealth is endogenous, and is, itself, a source of bias. Hum and Simpson (1994) recommend caution as there are many hidden pitfalls in the available methodology.

Sometimes, there are many available proxies for an unobservable variable but no theoretical reasons to choose among them. Goldberger (1974) discusses various inference methods for models with unobservables and multiple proxies. Some problems can be expressed as a simultaneous equation model, which can be estimated under sufficient identification assumptions. However, it is virtually impossible to check the validity of these assumptions in this context. Jöreskog (1973, 1978) suggests the identification of each unobservable variable with a common factor of its proxies. This identity is a strong assumption, which, if invalid, would generate biased estimators in the procedure. Furthermore, some dependence between observations is introduced through the factor analysis.

The first econometric works focused on linear regression models with proxy variables and measurement errors. In this context, only measurement errors on regressors affect Ordinary Least Squares (OLS) consistency. The parameters can be consistently estimated under identification assumptions using some observed instrument or specifying the probability distribution of observation errors. The literature is extensive (see Sargan (1958), Zellner (1970), Goldberger (1972), Robinson (1974), and the reviews of Aigner, Hsiao, Kapteeyn and Wansbeek (1984), Bowden and Turkington (1984), and Fuller (1987), among others). For non linear models, the existence of an instrument is not always sufficient to estimate their parameters consistently, although some

cases are tractable (see e.g., Amemiya, 1985a; Hausman, Newey, Ichimura, and Powell, 1991; Hausman, Newey and Powell, 1995; and Li and Hsiao, 2004). Measurement errors in the endogenous variables also affect consistency in such nonlinear models as binary choice models (Hausman, Avrevaya, and Scott-Morton, 1998), multinomial models (Hsiao and Sun, 1999), and count models (Li, Trivedi, and Guo, 2003). The usual procedure for estimating these models is to make a distributional assumption for the unobservable variables. The Expectation-Maximization (EM) algorithm (see Dempster, 1977) and the Simulated EM algorithm (see Wei and Tunner, 1990) are popular procedures for finding maximum likelihood estimators when some of the variables are unobserved but their probability distribution is postulated.

Despite advances in econometric theory, much can be done to enlarge the catalogue of techniques for estimating econometric models with unobservables. Further development in the enhancement of the alliance between economic theory and empirical analysis is worthwhile, and would be appreciated by practitioners (see e.g., Hum and Simpson, 1994).

The aim of this article is to present a robust method that we will call "worst-case (WC) estimation method" in order to estimate econometric models containing observable and unobservable variables. The estimation procedure guarantees the best parameter estimation in view of the worst-case values of the unobservable variables. The WC estimation method should be seen as a complement approach to standard techniques that postulate distributional assumptions for the unobservable variables. A cautious modeller should consider different estimation methods and balance the resulting estimates to determine a robust model. Robust modeling has been put forward in recent macroeconomic literature (see e.g. Hansen and Sargen, 2000). Under appropriate conditions, we prove consistency and asymptotic normality of WC estimators. We also discuss the relevance of these methods to reduce the adverse effect of the Lucas (1976) critique.

Worst-case techniques have been applied in game theory in the study of decision making in n-person conflicts (see e.g., Rosen, 1965). In a worst-case strategy, decision makers seek to minimize the maximum damage that their rival can inflict upon them. When the rival can be interpreted as nature, rather than another individual, the worst-case strategy seeks optimal responses in the worst-case value of uncertainty. Minimax principles have also been applied to different statistical problems, including such problem as the statistical efficiency of point estimators (see e.g., Lehmann, 1983, pp. 249-290), hypothesis tests for maximizing the minimum power when there is no uniformly most powerful test (see, Lehmann, 1986, Chapter 9), uniform bounds for the consistency of nonparametric density estimators (see e.g., Devroye, 1987), and optimal sampling designs from finite populations (see e.g., Gabler, 1990). As Huber (1994, pp. 59) points out, "the least favorable situation is safeguarding against, far from being unrealistically pessimistic, is more similar to actually observed error distributions than the normal model". Huber (1964) introduces a groundbreaking robust method of estimating location parameters for contaminated normal distributions, minimizing the maximal (worst-case) asymptotic variance that can happen over a neighborhood of the specified model (see also Huber, 1994, Chapter IV). The risk robustness of worst-case methods has also been appreciated in finance, with applications in portfolio management, see e.g. Rustem and Howe (2002).

In Section 2 of this paper we present the WC estimation method. Section 3 is devoted to the asymptotic properties of the WC estimators. Section 4 extends the

method to overidentified problems, presenting a worst-case Generalized Method of Moments (GMM). Because minimax problems usually turn out to be too unmanageable for closed solutions Section 5 addresses the numerical computation of WC estimators. In Section 6, we conduct a Monte Carlo simulation to study the finite sample behavior of WC estimators. Section 7 presents an illustration of the applicability of the method to Economics. Appendix A contains some instrumental results on minimax. Finally, proofs are placed in Appendix B.

2   The estimation method

A general framework encompassing most econometric estimators is the class of M-estimators, introduced by Huber (1964, 1967) as a generalization of Maximum Likelihood. These estimators are the minimizers of some loss function that depends on data. Let $\Theta \subset \mathbb{R}^K$ be a compact set of parameters and (X, Y) be a random vector of variables. Parameters θ^0 are defined as the minimizers of a loss function $Q(\theta) = E[g(X, Y, \theta)]$ on Θ, where g is a continuous function. Following the *analogy principle* (see Manski, 1994), given a sample $\{X_t, Y_t\}_{t=1}^T$ identically distributed as (X, Y), parameters θ^0 can be consistently estimated by minimizing $Q_T(\theta) = T^{-1} \sum_{t=1}^T g(X_t, Y_t, \theta)$ on Θ. This minimizer $\widehat{\theta}_T$ is known as M-estimator. Its asymptotic theory can be found in many econometric reviews (see e.g., Amemiya, 1985b; Bierens 1981, 1994; Wooldridge, 1994; and Pötscher and Prucha, 1991a, 1991b, 1997; among others). For a review of the statistical literature, see e.g., van der Vaart (1998) and van Geer (2000).

In this section we present the WC approach for determining and estimating optimal parameters of a model in a robust way against the worst-case value of the unobservable variables. Consider that the random vector Y is unobservable. Let $Q(\theta, y) = E[g(X, y, \theta)]$ denote the WC loss function, where $\theta \in \Theta$ and y is a vector of unobservables defined on $\mathbb{Y} \subset \mathbb{R}^S$. The worst-case strategy considers parameters θ^{wc} that solves the problem:

$$\min_{\theta \in \Theta} \max_{y \in \mathbb{Y}} Q(\theta, y). \tag{1}$$

These parameters θ^{wc} are those that best fit the available data in view of the unobservable variable Y. The WC strategy safeguards against the worst-case outcomes of the variable Y and makes no assumptions about the statistical nature of the unobservable.

Given the sample $\{X_t\}_{t=1}^T$, identically distributed as X, and the sample analog of $Q(\theta, y)$,

$$Q_T(\theta, y) = T^{-1} \sum_{t=1}^T g(X_t, y, \theta),$$

the WC estimator $\widehat{\theta}_T^{wc}$ of θ^{wc} is defined as the solution to:

$$\min_{\theta \in \Theta} \max_{y \in \mathbb{Y}} Q_T(\theta, y). \tag{2}$$

The worst-case criterion yields robust estimations in the sense that it protects against the distribution of the unobservable variables being concentrated on the "worst" state of nature. Despite there are no assumptions about the statistical nature of the unobservables, WC estimators have small asymptotic biases for standard regression models. Section 6 reports the bias of WC estimators with respect to the true parameters θ^0 for linear and non linear regression models. The Euclidean norm of these biases are 0.0006, and 0.004 respectively, showing the good performance of this approach.

As mentioned in the Introduction, there are approaches that consider a prior distribution F on $\mathbb{Y}$, when Y is unobservable. The parameter θ^F that minimizes $\int Q(\theta, y) F(dy)$ can be consistently estimated by minimizing its sample analogous $\int Q_T(\theta, y) F(dy)$. Here we discuss the reasons why risk averse modelers would prefer WC estimators. Let F_0 denote the true probability distribution for Y. Inequality $\max_{y \in \mathbb{Y}} Q(\theta^{wc}, y) \geq Q(\theta^{wc}, y)$ for all $y \in Y$ implies that

$$\max_{y \in \mathbb{Y}} Q(\theta^{wc}, y) \geq \int Q(\theta^{wc}, y) F_0(dy),$$

i.e., WC parameters guarantee an improvement of the true loss function, no matter which is the unknown F_0. This robustness is a key characteristics of WC methods. The true loss value associated with θ^{wc} is upper bounded, and the bound can be estimated by $\max_{y \in \mathbb{Y}} Q_T\left(\widehat{\theta}_T^{wc}, y\right)$. Therefore, robust WC estimators are a sensible alternative for testing economic theories.

F-integration methods do not have an upper bound as it is not guaranteed that $\int Q(\theta^F, y) F(dy) \geq \int Q(\theta^F, y) F_0(dy)$. However, when $F = F_0$, integrated methods outperform WC techniques as

$$\int Q(\theta^{wc}, y) F_0(dy) \geq \min_{\theta \in \Theta} \int Q(\theta, y) F_0(dy) = \int Q(\theta^F, y) F_0(dy).$$

Economists rarely have information on the probability of the unobservables, and the choice of F is typically a matter of convenience rather than an expression of actual knowledge. Hence, risk adverse modelers should avoid F-integration methods when F is only a postulated distribution.

In practice, the sets Θ and Y are determined in the light of economic theory and econometric literature. Without loss of generality, we assume that the parameters set is of the form $\Theta = \left\{\theta \in \mathbb{R}^K : h(\theta) \leq 0\right\}$, where h is a continuous vector function on Θ.

Analogously to nonlinear least squares methods, parametric constrains have no asymptotic effect if θ^{wc} is an interior point of Θ, i.e. $\theta^{wc} \in int\{\Theta\} = \left\{\theta \in \mathbb{R}^K : h(\theta) < 0\right\}$. However, if $\theta^{wc} \notin int\{\Theta\}$, the asymptotic distribution will be affected and this fact provides the basis for deriving asymptotic tests for the null hypothesis $H_0 : \left\{h_j(\theta^{wc}) = 0\right\}$.

In the next sections we examine the theoretical properties of $\left(\widehat{\theta}_T^{wc} - \theta^{wc}\right)$, such as existence, consistency, asymptotic normality and computation in a finite number

of steps. The compactness of Y is crucial for the arguments used in the proofs of these theoretical properties. If y has a multinomial distribution, $\mathbb{Y}$ is finite. But the compactness of $\mathbb{Y}$ often entails a truncation of Y. Nonetheless, choosing a large enough compact set, we can ensure that $\Pr(Y \notin \mathbb{Y}) < \varepsilon$ for an arbitrarily small $\varepsilon > 0$ (see e.g., Billingsley, 1968, Theorem 1.4). Additional notation should be introduced to derive the asymptotic distribution of WC estimators. Assume Q is continuous and $\mathbb{Y}$ is a nonempty compact set. For each $\theta \in \Theta$, there exists a set

$$\mathbb{Y}(\theta) = \left\{ y \in \mathbb{Y} : Q(\theta, y) = \max_{z \in \mathbb{Y}} Q(\theta, z) \right\}.$$

Therefore, $\mathbb{Y}(\theta^{wc})$ is the set of the worst-case unobservables. This set can be estimated by means of $\mathbb{Y}_T\left(\widehat{\theta}_T^{wc}\right)$, where

$$\mathbb{Y}_T(\theta) = \left\{ y \in \mathbb{Y} : Q_T(\theta, y) = \max_{z \in \mathbb{Y}} Q_T(\theta, z) \right\}.$$

Some properties of these sets are presented in Appendix A (see Lemma 8).

To obtain WC estimators, we are faced with the problem of solving a minimax continuous problem. Pioneering contributions to the study of minimax optimization have been made by Danskin (1967), Bram (1966), Rockafellar (1970), and Dem'yanov and Malozemov (1972). An algorithm for solving these problems is presented in Section 5.

Worst-case methods possess an additional interest for economic decision makers, as these methods can be used to reduce the damages derived from Lucas' critique. Lucas (1976) has pointed out that macro-econometric models cannot be used for policy analysis, if the implementation of the policy would change the conditional model on which the policy was based. The fact that agents have rational expectations over future policy actions, Lucas argued, turns this situation into a common problem. Control variables that are not affected by this problem are called super-exogenous. Consider an economic model where Y are variables controlled by the economic authority. If changes in the control variables Y affect the true parameters θ^0, we can use the worst-case parameter θ^{wc} which is relatively robust to changes in the controls Y. Therefore $\widehat{\theta}_T^{wc}$ could be a more stable tool for designing economic policies in the absence of super exogeneity. This approach can be used to design optimal economic policies in situations in which Lucas' (1976) critique would render impossible the success of traditional approaches.

3   Asymptotic properties of WC estimators

In this section we study the consistency and asymptotic normality of worst-case estimators. To prove consistency it is helpful to impose some regularity conditions.

A.1. For all T,
$$Q_T(\theta, y) - Q_T(\theta', y') = K_{\theta', y'}(\theta, y) - t_T(\theta, y)$$

where $K_{\theta', y'}(\theta, y)$ is a nonstochastic function, and $|t_T(\theta, y)| \to 0$ almost surely (in probability) when $T \to \infty$, uniformly in $\theta \in \Theta$ and $y \in \mathbb{Y}$.

$A.2$. For some $\theta^{wc} \in \Theta$ and $y^{wc} \in \mathbb{Y}(\theta^{wc})$, it is satisfied that, $\forall \varepsilon > 0$, $\exists \delta > 0$,

$$\inf_{\|\theta - \theta^{wc}\| \geq \varepsilon} \sup_{y \in \mathbb{Y}} K_{\theta^{wc}, y^{wc}} (\theta, y) > \delta.$$

The first assumption ensures that the objective function for WC estimation can be decomposed as the sum of a deterministic function and an asymptotically negligible stochastic term. Assumption $A.2$ requires that θ^{wc} solves problem $\inf_{\theta \in \Theta} \sup_{y \in \mathbb{Y}} K_{\theta^{wc}, y^{wc}} (\theta, y)$ uniquely in a neighborhood of θ^{wc} (uniqueness is an asymptotic identification requirement).

An alternative set of conditions to prove consistency can be given by means of the following tautology:

$$\begin{aligned}
Q_T(\theta, y) - Q_T(\theta^{wc}, z) &= K_{\theta^{wc}, z}(\theta, y) + t_T(\theta, y), \\
K_{\theta^{wc}, z}(\theta, y) &= Q(\theta, y) - Q(\theta^{wc}, z), \\
t_T(\theta, y) &= Q_T(\theta, y) - Q(\theta, y) + Q(\theta^{wc}, z) - Q_T(\theta^{wc}, z).
\end{aligned} \tag{3}$$

Then, it is sufficient for $A.1$ and $A.2$ (and therefore the consistency of $\widehat{\theta}_T^{wc}$) that $\theta^{wc} \in \Theta$ be a locally unique solution to (1), and that

$$\sup_{\theta \in \Theta} \sup_{y \in \mathbb{Y}} |Q_T(\theta, y) - Q(\theta, y)| \to 0,$$

almost surely (in probability). The uniform convergence of $Q_T(\theta, y) - Q(\theta, y)$ in $\Theta \times \mathbb{Y}$ can be checked using standard Uniform Laws of Large Numbers (ULLN). Dudley (1999, Section 6.6), van der Vaart and Wellner (1996, Section 2.4) and van Geer (2000) review ULLN literature for independent variables $\{X_t\}$. Davidson (1994, Chapter 21), Wooldridge (1994), and Pötscher and Prucha (1997, Chapter 5) review the econometric literature, including dependent data.

Theorem 1 *Consistency of WC estimators. Let $\widehat{\theta}_T^{wc} \in \Theta \subset \mathbb{R}^K$ be the solution to (2) with Q_T measurable for each $\theta \in \Theta$ and $y \in \mathbb{Y}$. Assuming $A.1$ and $A.2$,*

$$\widehat{\theta}_T^{wc} \to \theta^{wc}$$

almost surely (in probability).

The next results are necessary to derive the asymptotic distribution of WC estimators. First, we study the consistency of WC unobservables $\mathbb{Y}_T\left(\widehat{\theta}_T^{wc}\right)$ and consider the Hausdorff distance $d_H(A, B)$ between two non empty Euclidean sets A, B; i.e.,

$$d_H(A, B) = \max \left\{ \sup_{a \in B} d(a, B), \sup_{b \in B} d(A, b) \right\},$$

where $d(a, B) = \inf_{b \in B} \|a - b\|$ denotes the distance between the point a and the set B. For compact sets A and B, it is satisfied that $d_H(A, B) = 0$ if and only if $A = B$.

Note that $\mathbb{Y}_T\left(\widehat{\theta}_T^{wc}\right)$ and $\mathbb{Y}\left(\theta^{wc}\right)$ are compact when Q_T and Q satisfy Condition (i) of Lemma 8 (see Appendix A).

An additional condition ensures the consistency of WC unobservables $\mathbb{Y}_T\left(\widehat{\theta}_T^{wc}\right)$.

$A.3.\ \forall \varepsilon > 0,\ \exists \delta > 0,$
$$\inf_{\{y_1,y_2 \in \mathbb{Y}:\|y_1-y_2\|>\varepsilon\}}\ \inf_{\theta \in \Theta}\ K_{\theta,y_1}\left(\theta, y_2\right) > \delta.$$

Define $K_{\theta,y_1}\left(\theta, y_2\right) = Q\left(\theta, y_2\right) - Q\left(\theta, y_1\right)$, as in (3). A sufficient condition for $A.3$ is
$$\left|Q\left(\theta, y_2\right) - Q\left(\theta, y_1\right)\right| > r\left(\theta\right)\ f\left(\|y_2 - y_1\|\right),$$
where $f\left(x\right) > 0$ for all $x > 0$, and $\inf_{\theta \in \Theta} r\left(\theta\right) > 0$.

Proposition 2 *Consistency of WC unobservables.* *Under assumptions A.1, A.2 and A.3,* $d_H\left(\mathbb{Y}_T\left(\widehat{\theta}_T^{wc}\right), \mathbb{Y}\left(\theta^{wc}\right)\right) \rightarrow 0$, *almost surely (in probability).*

Similar behavior can be expected of WC multipliers $\{(\widehat{\mu}_i, \widehat{y}_i)\}_{i=1}^{\widehat{k}}$ associated with Problem (2) and $\widehat{\theta}_T^{wc}$ (see Appendix A, for definition and properties of WC multipliers). The next result gives sufficient conditions ensuring that $\{(\widehat{\mu}_i, \widehat{y}_i)\}_{i=1}^{\widehat{k}}$ converges almost surely to the WC multipliers $\{(\mu_i, y_i)\}_{i=1}^{k}$ associated with Problem (1) and θ^{wc}.

Proposition 3 *Consistency of WC multipliers. Under the assumptions in Theorem 9, A.1, A.2, A.3, if, in addition,* $\max_{\theta \in \Theta, y \in \mathbb{Y}}\left|Q_T\left(\theta, y\right) - Q\left(\theta, y\right)\right| \rightarrow 0$ *almost surely (in probability), then* $\{(\widehat{\mu}_i, \widehat{y}_i)\}_{i=1}^{\widehat{k}}$ *converges to* $\{(\mu_i, y_i)\}_{i=1}^{k}$ *almost surely (in probability).*

Next we obtain the asymptotic distribution of WC estimators, under the following assumptions:

$B.1.\ \theta^{wc} \in int\left\{\Theta\right\}$ solves (1), and $\widehat{\theta}_T^{wc} \rightarrow_p \theta^{wc}$,

$B.2.\ \{(\widehat{\mu}_i, \widehat{y}_i)\}_{i=1}^{\widehat{k}} \rightarrow_p \{(\mu_i, y_i)\}_{i=1}^{k}$,

$B.3.$ for all T, $Q_T\left(\theta, y\right)$ is $C^{2,1}$ almost surely, $\mathbb{Y} \subset \mathbb{R}^S$ and $\Theta \subset \mathbb{R}^K$ are nonempty compact sets, and
$$\sqrt{T}\frac{\partial Q_T\left(\theta^{wc}, y\right)}{\partial \theta} \rightarrow_d Z\left(y\right)$$

uniformly on $C\left(\mathbb{Y}\right)$, where Z is a second order Gaussian process, with zero mean and covariance $R\left(y_1, y_2\right) = E\left[Z\left(y_1\right)Z\left(y_2\right)'\right]$,

$B.4$. for any sequence $\tilde{\theta}_T \to_p \theta^{wc}$,

$$\sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \frac{\partial^2 Q_T \left(\tilde{\theta}_T, \widehat{y}_i\right)}{\partial\theta\partial\theta'} \to_p B := \sum_{i=1}^{k} \mu_i \frac{\partial^2 Q \left(\theta^{wc}, y_i\right)}{\partial\theta\partial\theta'},$$

where B is a nonsingular deterministic real matrix.

Theorem 4 *Asymptotic Normality.* Let $\widehat{\theta}_T^{wc}$ be the solution to (2). Assume $B.1$, $B.2$, $B.3$ and $B.4$. Then, $\sqrt{T}\left(\widehat{\theta}_T^{wc} - \theta^{wc}\right) \to_d N\left(0, B^{-1}AB^{-1}\right)$, where $A = \sum_{i=1}^{k}\sum_{j=1}^{k} \mu_i\mu_j R\left(y_i, y_j\right)$ is a positive definite real matrix.

Consistency of WC estimators and multipliers (considered in Assumptions $B.1$ and $B.2$) can be proven using Theorem 1 and Propositions 2 and 3. Assumption $B.3$ can be established applying a standard functional central limit theorem for empirical processes. These central limit theorems require weak convergence of finite dimensional projections and a *uniform tightness* Condition. For an introduction to this topic, see Billingsley's (1968) classical monograph, Wichura (1969), and Bickel and Wichura (1971). Pollard (1989, 1990), Dudley (1999) and van der Vaart and Wellner (1996) review a different approach, particularly fruitful under independence assumptions. Assumption $B.4$ can be derived from the uniform consistency condition

$$\left\| \frac{\partial^2 Q_T\left(\theta, y\right)}{\partial\theta\partial\theta'} - \frac{\partial^2 Q\left(\theta, y\right)}{\partial\theta\partial\theta'} \right\| \to_p 0$$

uniformly on $C\left(\mathbb{Y} \times \Theta\right)$, which requires a ULLN. Assumption $B.4$ can also be established applying the following result:

Proposition 5 *Sufficient conditions for $B.4$ are:*

C.1. $B_T = \sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \partial^2 Q_T\left(\theta^{wc}, \widehat{y}_i\right) \Big/ \partial\theta\partial\theta' \to_p B$, *and*

C.2. $E\left[\sup_{\|\theta - \theta^{wc}\| \leq \delta} \sup_{y \in \mathbb{Y}} \left\| \partial^2 Q_T\left(\theta, y\right) \Big/ \partial\theta\partial\theta' - \partial^2 Q_T\left(\theta^{wc}, y\right) \Big/ \partial\theta\partial\theta' \right\| \right] \xrightarrow[\delta\downarrow 0]{} 0$, *for all T.*

Condition $C.1$ follows from $B.2$, whenever $\partial^2 Q_T\left(\theta^{wc}, y\right) \Big/ \partial\theta\partial\theta' \to \partial^2 Q\left(\theta^{wc}, y\right) \Big/ \partial\theta\partial\theta'$ uniformly on $C\left(\mathbb{Y}\right)$. For Condition $C.2$ it is sufficient that

$$\left\| \frac{\partial^2 Q_T\left(\theta, y\right)}{\partial\theta\partial\theta'} - \frac{\partial^2 Q_T\left(\theta^{wc}, y\right)}{\partial\theta\partial\theta'} \right\| \leq f_T\left(y\right) \|\theta - \theta^{wc}\|^{\alpha} \tag{4}$$

for some $\alpha \in (0, 1)$, and $E\left[\sup_{y \in \mathbb{Y}} |f_T\left(y\right)|\right] < \infty$. For (4), it suffices that the elements in $\partial^2 g\left(\theta, y\right) / \partial\theta\partial\theta'$ satisfy a Lipschitz condition.

Often,

$$R\left(y_1, y_2\right) = \lim_{T \to \infty} \frac{1}{T} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} E\left[\frac{\partial g\left(X_{t_1}, y_i, \theta\right)}{\partial \theta} \frac{\partial g\left(X_{t_2}, y_j, \theta\right)}{\partial \theta'}\right].$$

Therefore, if $\{X_t\}$ are independently distributed, A can be estimated by

$$\widehat{A}_T = \sum_{i=1}^{\widehat{k}} \sum_{j=1}^{\widehat{k}} \widehat{\mu}_i \widehat{\mu}_j \left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial g\left(X_t, \widehat{y}_i, \widehat{\theta}_T^{wc}\right)}{\partial \theta} \frac{\partial g\left(X_t, \widehat{y}_j, \widehat{\theta}_T^{wc}\right)}{\partial \theta'}\right),$$

and B by

$$\widehat{B}_T = \sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \frac{\partial^2 Q_T\left(\widehat{\theta}_T^{wc}, \widehat{y}_i\right)}{\partial \theta \partial \theta'} = \sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 g\left(X_t, \widehat{y}_i, \widehat{\theta}_T^{wc}\right)}{\partial \theta \partial \theta'}\right).$$

Analogously, we can establish sufficient conditions for the asymptotic normality of WC estimators when $\theta^{wc} \notin int\left\{\Theta\right\}$ (i.e. there exist some p such that $h_j\left(\theta^{wc}\right) = 0$ for $j = 1, ..., p$). In particular, we obtain the asymptotic distribution of $\sqrt{T}\left(\widehat{\theta}_T^{wc} - \theta^{wc}, \widehat{\lambda}_T^{wc}\right)$, which allows us to derive asymptotic parametric tests. The proof of asymptotic normality is similar to that of Theorem 4; however, we should slightly modify Assumption $B.1$ as follows:

$B.1'$. $\theta^{wc} \in \Theta$ solves (1), satisfying that $h_j\left(\theta^{wc}\right) = 0$ for $j = 1, ..., p$, and $h_j\left(\theta^{wc}\right) < 0$ for $j = p+1, ..., P$, where $\left\{\partial h_j\left(\theta^{wc}\right)/\partial \theta\right\}_{j=1}^{p}$ are linearly independent. Also, $\widehat{\theta}_T^{wc} \to_p \theta^{wc}$.

Theorem 6 *Asymptotic Normality of constrained WC estimators. Let $\widehat{\theta}_T^{wc}$ be the solution to (2) with Lagrange multipliers $\widehat{\lambda}^{wc}$. Assume $B.1'$, $B.2$, $B.3$ and $B.4$. Then,*

$$\sqrt{T}\left(\begin{array}{c} \widehat{\theta}_T^{wc} - \theta^{wc} \\ \\ \widehat{\lambda}^{wc} \end{array}\right) \to_d N\left(0, V\right),$$

where,

$$V = \left(\begin{array}{cc} B & H' \\ \\ H & 0 \end{array}\right)^{-1} \left(\begin{array}{cc} A & 0 \\ \\ 0 & 0 \end{array}\right) \left(\begin{array}{cc} B & H' \\ \\ H & 0 \end{array}\right)^{-1},$$

with $H = \nabla_\theta H_p\left(\theta^{wc}\right)$, $H_p\left(\theta\right) = \left(h_1\left(\theta\right), ..., h_p\left(\theta\right)\right)'$, and matrices A and B as defined in Theorem 4.

9

Consider

$$
\begin{pmatrix} C_{11} & C'_{12} \\ C_{12} & C_{22} \end{pmatrix} = \begin{pmatrix} B & H' \\ H & 0 \end{pmatrix}^{-1}.
$$

We can express the asymptotic covariance matrix as,

$$
V = \begin{pmatrix} V_{11} & V'_{12} \\ V_{12} & V_{22} \end{pmatrix} = \begin{pmatrix} C_{11}AC_{11} & C'_{12}AC_{11} \\ C_{11}AC_{12} & C'_{12}AC_{12} \end{pmatrix}.
$$

The explicit form of this matrix can be obtained applying standard formulae for the inverse of a partitioned matrix,

$$
\begin{aligned}
C_{11} &= B^{-1} - B^{-1}H' \left(HB^{-1}H' \right)^{-1} HB^{-1}, \\
C_{12} &= \left(HB^{-1}H' \right)^{-1} HB^{-1}.
\end{aligned}
$$

If $A = B$, we can simplify V to,

$$
\begin{pmatrix} V_{11} & V'_{12} \\ V_{12} & V_{22} \end{pmatrix} = \begin{pmatrix} B^{-1} \left(I - H' \left(HB^{-1}H' \right)^{-1} HB^{-1} \right) & 0 \\ 0 & \left(HB^{-1}H' \right)^{-1} \end{pmatrix}.
$$

The unrestricted WC estimate, by contrast, has an asymptotic covariance matrix B^{-1}, and thus is generally less efficient than the constrained WC estimator (as $B^{-1}H' \left(HB^{-1}H' \right)^{-1} HB^{-1}$ is nonnegative definite). It means that by incorporating valid restrictions we cannot reduce efficiency, but generally improve it.

Theorem 6 can be used to test Lagrange multiplier hypotheses. For example, the statistic for testing $H_0 : h_j \left(\theta^{wc} \right) = 0$ for $j = 1, ..., p$, is $\Upsilon_T := T \widetilde{\lambda}' \widehat{V}_{22}^{-1} \widehat{\lambda} \to_d \chi^2_{K-p}$, where $\widehat{V}_{22} \to_p V_{22} > 0$. Other asymptotic tests, such as Wold type tests and generalized likelihood ratio tests, can be derived in a similar way, using $\sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \nabla_\theta Q_T \left(\theta, \widehat{y}_i \right)$ as a score function.

4 Worst-Case estimation for overidentified models

Hansen's (1982, 1985) GMM for overidentified models consider θ^0 as the minimizer of a quadratic loss function $Q \left(\theta \right) = E \left[g \left(X, Y, \theta \right) \right]' W E \left[g \left(X, Y, \theta \right) \right]$ on Θ, where W is a positive definite matrix. Following the analog principle, the parameters are consistently estimated by the minimizer of

$$
Q_T \left(\theta \right) = \left(T^{-1} \sum_{t=1}^{T} g \left(X_t, Y_t, \theta \right) \right)' W_T \left(T^{-1} \sum_{t=1}^{T} g \left(X_t, Y_t, \theta \right) \right),
$$

where $W_T \to_p W$. A review of the literature can be found, e.g. in Hall (1993), Ogaki (1993), Newey (1993), Newey and McFadden (1994) and Wooldridge (1994), among others.

This section show how the worst-case approach is embedded in the GMM framework. Assume that Y is unobserved and consider the loss function

$$Q(\theta, y) = E\left[g(X, y, \theta)\right]' W E\left[g(X, y, \theta)\right]$$

on $\Theta \times \mathbb{Y}$. We define θ^{wc} as the solution of $\min_{\theta \in \Theta} \max_{y \in \mathbb{Y}} Q(\theta, y)$. Given the sample data $\{X_t\}_{t=1}^{T}$ and the sample analog

$$Q_T(\theta, y) = \left(T^{-1}\sum_{t=1}^{T} g(X_t, y, \theta)\right)' W_T \left(T^{-1}\sum_{t=1}^{T} g(X_t, y, \theta)\right), \qquad (5)$$

where $W_T \to_p W$ almost surely (in probability), the WC GMM estimator $\widehat{\theta}_T^{wc}$ of θ^{wc} is defined as the solution to $\min_{\theta \in \Theta} \max_{y \in \mathbb{Y}} Q_T(\theta, y)$.

Consistency results derived in the previous sections are valid to this extension (see Theorem 1 and Proposition 2). However, to prove asymptotic normality, we should consider the following assumptions:

$D.1.$ $\theta^{wc} \in int\{\Theta\}$ solve $\min_{\theta \in \Theta} \max_{y \in \mathbb{Y}} Q(\theta, y)$, where

$$Q(\theta, y) = E\left[g(X, y, \theta)\right]' W E\left[g(X, y, \theta)\right],$$

W is positive definite, and $\widehat{\theta}_T^{wc} \to_p \theta^{wc}$,

$D.2.$ $\{(\widehat{\mu}_i, \widehat{y}_i)\}_{i=1}^{\widehat{k}} \to_p \{(\mu_i, y_i)\}_{i=1}^{k}$,

$D.3.$ $g(X, \theta, y) \in C^{1,1}(\Theta \times \mathbb{Y})$ almost surely, $\mathbb{Y} \subset \mathbb{R}^S$ and $\Theta \subset \mathbb{R}^K$ are nonempty compact sets, and

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g(X_t, y, \theta^{wc}) \to_d G(y)$$

uniformly on $C(\mathbb{Y})$, where G is a second order Gaussian process, with zero mean and covariance $R(y_1, y_2) = E\left[G(y_1) G(y_2)'\right]$,

$D.4.$
$$T^{-1}\sum_{t=1}^{T}\frac{\partial}{\partial \theta}g(X_t, y, \theta) \to_p S(y, \theta) := E\left[\frac{\partial}{\partial \theta}g(X, y, \theta)\right]$$

uniformly on $C(\mathbb{Y} \times \Theta)$.

Theorem 7 *Asymptotic Normality of WC GMM.* Let $\widehat{\theta}_T^{wc}$ be the solution to (2), and Q_T be given by (5). Assume D.1, D.2, D.3, and D.4. Then,

$$\sqrt{T}\left(\widehat{\theta}_T^{wc} - \theta^{wc}\right) \to_d N\left(0, E^{-1}DE^{-1}\right),$$

where

$$
\begin{aligned}
D &= \sum_{i=1}^{k}\sum_{j=1}^{k} \mu_i\mu_j\, S\left(y_i, \theta^{wc}\right)\, W\, R\left(y_i, y_j\right)\, W'\, S\left(y_j, \theta^{wc}\right)', \\
E &= \sum_{i=1}^{k} \mu_i\, S\left(y_i, \theta^{wc}\right)\, W\, S\left(y_i, \theta^{wc}\right)'.
\end{aligned}
$$

The asymptotic variance of WC GMM estimators is more complex than the one of the classical GMM estimators. If $R\left(y_i, y_j\right) = S\left(y_i, \theta^{wc}\right) S\left(y_j, \theta^{wc}\right)'$, the asymptotic variance is $E^{-1}DE^{-1} = I$. In the WC context, if $k > 1$ it is not straightforward to ensure that $D = E$ by an appropriate choice of W.

The asymptotic distribution for WC GMM constrained estimators can be derived analogously to Theorem 6.

5 Computational issues

Computing WC estimators involves solving a minimax continuous problem. We use the global optimization algorithm developed by Shimizu and Aiyoshi (1980), see also Žaković and Rustem (2003). They consider an algorithm for solving semi-infinite programming problems, as any continuous minimax problem of the form $\min_{\theta\in\Theta}\max_{y\in\mathbb{Y}} Q_T\left(\theta, y\right)$ can be written as $\min_{\theta\in\Theta,\, \rho\in\mathbb{R}}\{\rho : \max_{y\in\mathbb{Y}} Q\left(\theta, y\right) \leq \rho\}$, which is equivalent to the semi-infinite problem:

$$\min_{\theta\in\Theta,\, \rho\in\mathbb{R}} \quad \rho \tag{6}$$

$$\text{s.t.} \qquad Q_T\left(\theta, y\right) \leq \rho, \text{ for all } y \in \mathbb{Y}.$$

This algorithm uses a global optimization approach with respect to $y \in \mathbb{Y}$ and cutting planes to reduce the feasible region when constraints violation is encountered. In particular, the l-th iteration of this algorithm consists of solving the problem:

$$\min_{\theta^{l+1}\in\Theta,\, \rho^{l+1}\in\mathbb{R}} \left\{\rho^{l+1}:\; Q_T\left(\theta^{l+1}, y_i\right) \leq \rho^{l+1}, \qquad i = 1, ..., k_l\right\}, \tag{7}$$

given $\{y_i\}_{i=1}^{k_l} \subset \mathbb{Y}\left(\theta^l\right)$. Next we check if the solution is feasible up to an arbitrary positive tolerance ε. If

$$\max_{y\in\mathbb{Y}} Q_T\left(\theta^{l+1}, y\right) > \rho^{l+1} + \varepsilon,$$

iterate, else if $\max_{y \in \mathbb{Y}} Q_T \left(\theta^{l+1}, y \right) \leq \rho^{l+1} + \varepsilon$, terminate and $\widehat{\theta}_T^{wc} = \theta^{l+1}$ is a solution of the minimax problem. This algorithm terminates in a finite number of iterations (see e.g., Shimizu and Aiyoshi, 1980, Theorem 3). Under convexity assumptions of Problem (6), the Lagrange multipliers $\{\mu_i\}_{i=1}^k$ associated with the last iteration of Problem (7) are the coefficients $\{\mu_i\}_{i=1}^k$ in Theorem 9.

The global optimization approach is essential to guarantee the robustness property of the solution of minimax problems because one of the crucial steps in solving the semi-infinite problem is to find $\{y_i\}_{i=1}^{k_l} \subset \mathbb{Y}_T \left(\theta^l \right)$ for all $\theta^l \in \Theta$ by computing the global maximizers in the program $\max_{y \in \mathbb{Y}} Q_T \left(\theta^l, y \right)$. In global optimization algorithms, all candidates for local maximizers must usually be bracketed by a comparison of function values $Q_T \left(\theta^l, y \right)$ on a sufficiently dense finite subset of $\mathbb{Y}$. To reduce the cost of computing global optima, it is recommended that the domains Θ and $\mathbb{Y}$ be restricted as much as possible given the information available. Computing time can be saved by a parallel computation of the maximizers.

Problem (6) can be efficiently computed using standard nonlinear programming packages. In Sections 6 and 7, we use the MATLAB subroutine *fseming* (http://www.mathworks.com).

The surveys in Hettich and Kortanek (1993) and Reemtsen and Görner (1998) provide an excellent introduction to semi-infinite programming algorithms. The monograph of Rustem and Howe (2002) is focused on the computation of minimax problems.

6 Monte Carlo study of finite sample behavior

A small Monte Carlo study was conducted in order to study the finite sample performance of worst-case estimates. We first consider the linear regression model (i) $y_t = \alpha x_{1t} + \beta x_{2t} + u_t$, where (α, β) denote the 3-dimensional parameter vector, and $x_{1t} \sim N(0,1)$, $x_{2t} \sim U([-3,-1])$, and $u_t \sim N(0,1)$ are identically and independently distributed random variables, for all $t = 1, ..., T$; where $N(0,1)$ denotes the standard normal distribution and $U([a,b])$ the uniform distribution on the interval $[a,b]$. Assuming $\alpha^0 = -2$ and $\beta^0 = 2$, the experiment was carried out for $T = 30$ and $T = 40$.

Consider the problem of estimating the WC estimators for (α, β) assuming that x_2 is unobservable. Following the nonlinear least-square approach, we define

$$Q_T \left(\alpha, \beta, x_2 \right) = T^{-1} \sum_{t=1}^{T} \left(y_t - \left(\alpha x_{1t} + \beta x_{2t} \right) \right)^2,$$

and the worst-case problem as

$$\min_{\alpha \in [-3,-1], \beta \in [1,3]} \max_{x_2 \in [-3,-1]} Q_T \left(\alpha, \beta, x_2 \right).$$

In order to illustrate the accuracy of the asymptotic distribution, we perform a Monte Carlo with $N = 400$ realizations. Table 1 reports expectation, variance and

absolute bias with respect to true parameters of WC estimators, $\widehat{\alpha}^{wc}$ and $\widehat{\beta}^{wc}$, using $T = 30$ and $T = 40$. The Euclidean norm of the WC bias with respect to the true parameters is equal to 0.00152 for $T = 30$ and 0.0006 for $T = 40$.

	$T = 30$	$T = 40$
$E\left[\widehat{\alpha}^{wc}\right]$	-2.0373	-2.0018
$V\left[\widehat{\alpha}^{wc}\right]$	0.1401	0.1175
$\left|E\left[\widehat{\alpha}^{wc}\right] - \alpha^0\right|$	0.0013	$3.45E - 06$
$E\left[\widehat{\beta}^{wc}\right]$	1.9713	1.9744
$V\left[\widehat{\beta}^{wc}\right]$	0.0202	0.0177
$\left|E\left[\widehat{\beta}^{wc}\right] - \beta^0\right|$	0.0008	0.0006

Table 1. Finite sample results for Model (i)

Figure 1 displays the normal probability plot for $\widehat{\alpha}^{wc}$ and $\widehat{\beta}^{wc}$, respectively, for $T = 30$.
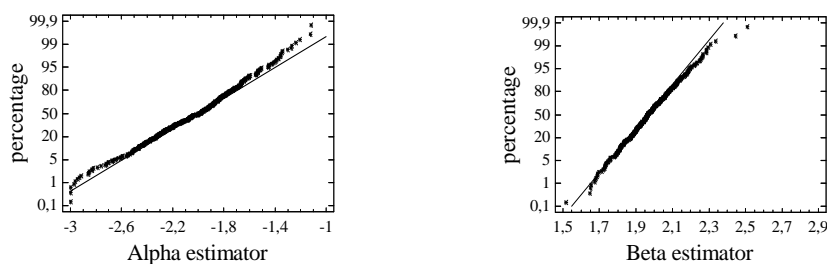

Figure 1. Normal Probability Plot for WC estimators model (i), with $T = 30$.

Note that linear regression models with an intercept δ^0 are not identified for the WC approach because once the WC variable is determined, the parameters are estimated by OLS given a fixed value of the unobservable, which is collinear with the

constant regressor. Nonetheless, we can still apply WC strategies using a linear model without intercept and mean-centered variables. An indicative value of the intercept can be given by $\widehat{\delta} = \overline{y} - \widehat{\alpha}^{wc}\overline{x}_1 - \widehat{\beta}^{wc}\overline{x}_2$, where $\overline{y}$ and $\overline{x}_1$ are the sample averages of the observables and $\overline{x}_2$ is an average of computed worst-case values.

A nonlinear regression model (ii) $y_t = \exp(\alpha + \beta x_{1t} + \gamma x_{2t}) + u_t$, was also simulated, where (α, β, γ) denote the 3-dimensional parameter vector, and $x_{1t} \sim N(0,1)$, $x_{2t} \sim U([-2,2])$, and $u_t \sim N(0,1)$ are identically and independent distributed random variables, for all $t = 1, ..., T$. Assuming $\alpha^0 = 0.9$, $\beta^0 = 0.75$, and $\gamma^0 = 0.2$, a Monte Carlo with $N = 300$ was carried out for $T = 30$ and $T = 40$. Table 2 reports expectation, variance and absolute bias with respect to true parameters of WC estimators $\widehat{\alpha}^{wc}$, $\widehat{\beta}^{wc}$ and $\widehat{\gamma}^{wc}$, when x_2 is unobservable. The Euclidean norm of the WC bias with respect to the true parameters is 0.004 for $T = 30$ and $T = 40$. The WC absolute bias is larger for $\widehat{\gamma}^{wc}$ than for the estimators $\widehat{\alpha}^{wc}$ and $\widehat{\beta}^{wc}$ related to observable variables, by contrast with model (i) where the parameter associated to the unobservable has lower bias.

	$T = 30$	$T = 40$
$E\left[\widehat{\alpha}^{wc}\right]$	1.0084	1.0054
$V\left[\widehat{\alpha}^{wc}\right]$	0.0064	0.0048
$\left\|E\left[\widehat{\alpha}^{wc}\right] - \alpha^0\right\|$	0.0117	0.0111
$E\left[\widehat{\beta}^{wc}\right]$	0.7458	0.7443
$V\left[\widehat{\beta}^{wc}\right]$	0.0036	0.0033
$\left\|E\left[\widehat{\beta}^{wc}\right] - \beta^0\right\|$	$E-05$	$E-05$
$E\left[\widehat{\gamma}^{wc}\right]$	0.002	0.0008
$V\left[\widehat{\gamma}^{wc}\right]$	0.0003	0.0003
$\left\|E\left[\widehat{\gamma}^{wc}\right] - \gamma^0\right\|$	0.0391	0.0396

Table 2. Finite sample results for Model (ii)

Figure 2 shows the normal probability plot for $\widehat{\alpha}^{wc}$, $\widehat{\beta}^{wc}$ and $\widehat{\gamma}^{wc}$, respectively, for $T = 30$.
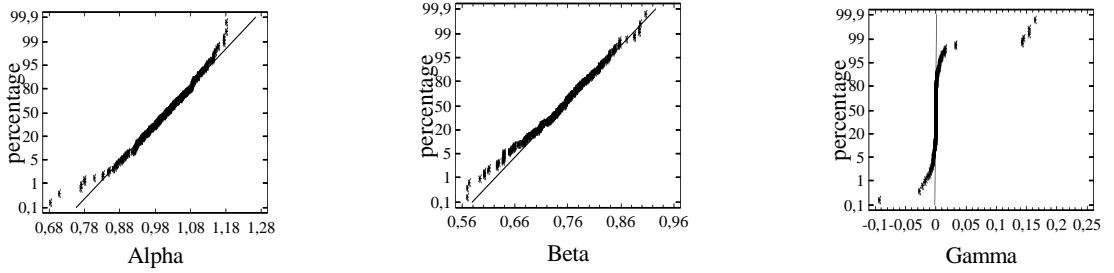
Figure 2. Normal Probability Plot for WC estimators model (ii), with $T = 30$.

Figures 1 and 2 show that the normal approximation is satisfactory for $T = 30$, although there is room for improvement in the tails. Second order asymptotic methods, such as Edgeworth expansion-based corrections or some resampling methods (e.g., bootstrap and subsampling), seem to provide interesting approaches for WC inferences with small samples. But for large samples they may not be worth emphasizing over first order weak asymptotic approximations, as in classical M-estimation.

7   An economic example

In this section we present an empirical application to illustrate the economic interest of the presented method. Consider the problem given in Hansen and Singleton (1982). Assume that a representative agent decides about consumption and investment, solving the dynamic optimization problem:

$$\max_{\{c_t, w_t\}_{t=0}^{\infty}} \left\{ \sum_{t=0}^{\infty} \theta_1^r E\left[u\left(c_t\right) | \mathcal{F}_0\right] \quad : c_t + \sum_{j=1}^{N} p_{j,t} q_{j,t} \leq \sum_{j=1}^{N} r_{j,t} q_{j,t-m_j} + w_t \right\},$$

where c_t denotes consumption, w_t denotes labor income, and q_t is a portfolio of N assets with respective maturities m_j, with spot price $p_{j,t}$ and payoff $r_{j,t}$ by stock at time $t - m_j$. The utility function u satisfies $u_c > 0$, $u_{cc} < 0$, and $\theta_1 \in (0,1)$ is the subjective discount factor. Furthermore, $\mathcal{F}_0$ is the information set available at time t. The solution to this problem satisfies,

$$
\begin{aligned}
p_{j,t} u'\left(c_t\right) &= \theta_1^{m_j} E\left[r_{j,t+m_j} u'\left(c_{t+m_j}\right) | \mathcal{F}_t\right] \Leftrightarrow \\
0 &= E\left[\left(\theta_1^{m_j} \frac{u'\left(c_{t+m_j}\right)}{u'\left(c_t\right)} \frac{r_{j,t+m_j}}{p_{j,t}} - 1\right) | \mathcal{F}_t\right] \qquad j = 1, ..., N.
\end{aligned}
$$

See Hansen and Singleton (1982) for details.

Assuming that $u(c) = c^{1-\theta_2}/1 - \theta_2$ where $\theta_2 > 0$, $\theta_2 \neq 1$ is the coefficient of relative risk aversion, and $u(c) = \log c$ when $\theta_2 = 1$, then

$$E\left[\left(\theta_1^{m_j} \left(\frac{c_{t+m_j}}{c_t}\right)^{1-\theta_2} \frac{r_{j,t+m_j}}{p_{j,t}} - 1\right) |\mathcal{F}_t\right] = 0, \qquad j = 1, ..., N.$$

Therefore, for any set Z_t known in t, the actual $\theta = (\theta_1, \theta_2)'$ satisfies,

$$E\left[\left(\theta_1^{m_j} \left(\frac{c_{t+m_j}}{c_t}\right)^{1-\theta_2} \frac{r_{j,t+m_j}}{p_{j,t}} - 1\right) Z_t\right] = 0, \qquad j = 1, ..., N.$$

Following Hansen and Singleton (1982), this expression can be used to estimate by GMM when all the required information is available. In this case, they considered that a subset of $r_{j,t+m_j}/p_{j,t}$ is observed for a subgroup of the N assets. Unfortunately, the GMM methodology cannot be applied if some of these variables have not been observed. Often, the spot price of an asset is not observed in the sampled range, but traders have an idea about its variation rank. For example, this happens when a new asset j is introduced in the market. The worst-case approach presented can be an useful tool to obtain an indicative value of model parameters.

Then, assuming that $p_{j,t}$ are not observed but take values in the range $[115, 180]$, we consider the worst-case GMM estimation associated with the moment conditions

$$E\left[g(X_t, y, \theta)\right] = E\left[\left(\theta_1^{m_j} \left(\frac{c_{t+m_j}}{c_t}\right)^{1-\theta_2} \frac{r_{j,t+m_j}}{y} - 1\right) Z_t\right] = 0,$$

with $X_t = \left(c_t, c_{t+m_j}, r_{j,t+m_j}, Z_t\right)$. In particular we solve

$$\min_{0 \leq \theta_1 \leq 1, \ 1 \leq \theta_2 \leq 30} \ \max_{115 \leq y \leq 180} \ \left(\frac{1}{T-2} \sum_{t=2}^{T-1} g(X_t, y, \theta)\right) W_T \left(\frac{1}{T-2} \sum_{t=2}^{T-1} g(X_t, y, \theta)\right),$$

with W_T the identity matrix, $m_j = 1$, $Z_t = (r_t, r_{t-1})'$, and $X_t = (c_t, c_{t+1}, r_{j,t+1}, Z_t')'$, for $t = 2, ..., T-1$.

Taking the equally weighted return on IBM stocks listed on the New York Stock Exchange (see http://www. princeton.edu/~data/datalib/timeser.html) and the real personal consumption expenditures of durable goods from the Federal Reserve (see http://www.economagic.com/ fedstl.htm#CPI) during 1986-1987, the worst-case parameters estimates obtained using the described procedure are $\widehat{\theta}_1^{wc} = 0.85$ and $\widehat{\theta}_2^{wc} = 1.25$, the maximum optimum in $\mathbb{Y}_T\left(\widehat{\theta}\right)$ is $\widehat{y}_1 = 155$, and the associated Lagrange's multiplier is $\widehat{\mu}_1 = -1$.

Note that when a tax is about to be introduced in the financial market such that the spot prices $p_{j,t}$ will be modified, the estimated WC parameters are more robust to price changes than are the ordinary parameters estimated by GMM. Therefore, if

the tax decision is based on the estimated model, an analysis based on WC modelling is less sensitive to Lucas' (1976) critique.

APPENDIX A: MINIMAX

The following result summarizes some properties of sets $\mathbb{Y}_T(\theta)$ and $\mathbb{Y}(\theta)$, that will be used to establish consistency. Below we present the first order conditions for minimax problems, that are used to establish the asymptotic normality.

Lemma 8 *Let $\Theta \subset \mathbb{R}^K$ and $\mathbb{Y} \subset \mathbb{R}^S$ be non empty compact sets and $Q \in C(\Theta \times \mathbb{Y})$. Then there exist $\theta^{wc} \in \Theta$ and $y^{wc} \in \mathbb{Y}$ such that*

$$Q(\theta^{wc}, y^{wc}) = \min_{\theta \in \Theta} \max_{y \in \mathbb{Y}} Q(\theta, y).$$

Furthermore, the set $\mathbb{Y}(\theta)$ satisfies the following properties: (i) If $Q(\theta, \cdot)$ is concave on $\mathbb{Y}$ for each $\theta \in \Theta$ and $\mathbb{Y}$ is convex, then the correspondence $\mathbb{Y}(\theta)$ is upper hemi-continuous and takes values that are non empty compact convex sets. (ii) If $Q(\theta, \cdot)$ is strictly convex for each $\theta \in \Theta$ and $\mathbb{Y}$ is convex, then $\mathbb{Y}(\theta)$ is a continuous function. (iii) The same properties can be applied to $\mathbb{Y}_T(\theta)$ when $Q_T(\theta, \cdot)$ is (strictly) concave for each $\theta \in \Theta$.

The minimax existence follows from a standard application of the Weierstrass Theorem and Berge's (1963) Maximum Theorem. The result is a consequence of the Maximum Theorem under convexity assumptions (see Sundaram, 1996, Theorem 9.17, pp. 237-238) that is a consequence of Berge's (1963) Theorem.

The following Theorem provides first order necessary conditions for the solution to (1). It is usually credited to Schmitendorf (1977) (see also Shimizu and Aiyoshi, 1980, Theorem 1). Nonetheless, first order conditions for minimax optima have been previously considered in the Russian literature, and translated into English before 1977 (see e.g., Dem'yanov and Malozemov, 1972). Also, there exist sufficient conditions for a point satisfying the first order conditions to be a minimax optima (e.g., Bector and Bhatia, 1985).

Theorem 9 *First order conditions for Minimax problems. Let $Q : \mathbb{R}^K \times \mathbb{R}^S \to \mathbb{R}$ be C^1, $\mathbb{Y} \subset \mathbb{R}^S$ be a nonempty compact set, and $\Theta = \{\theta \in \mathbb{R}^K : h(\theta) \leq 0\}$ where $h : \mathbb{R}^K \to \mathbb{R}^P$ are C^1. Let θ^{wc} denote the solution to $\min_{\theta \in \Theta} \max_{y \in \mathbb{Y}} Q(\theta, y)$. If vectors $\{\nabla_\theta h_j(\theta^{wc}) : h_j(\theta^{wc}) = 0\}$ are linearly independent, then there exist a positive integer k, vectors $y_i \in \mathbb{Y}(\theta^{wc})$, and multipliers $\mu_i \geq 0$ for $i = 1, ..., k$, with $\sum_{i=1}^{k} \mu_i = 1$, and $\lambda_j \geq 0$ for $j = 1, ..., p$ such that*

$$\sum_{i=1}^{k} \mu_i \nabla_\theta Q(\theta^{wc}, y_i) + \sum_{j=1}^{p} \lambda_j \nabla_\theta h_j(\theta^{wc}) = 0, \tag{8}$$

$$\sum_{j=1}^{p} \lambda_j h_j(\theta^{wc}) = 0, \tag{9}$$

with $1 \leq k + p \leq K + 1$, where p is the number of nonzero λ_j. If $\theta^{wc} \in int\{\Theta\}$, Equations (8) and (9) simplify to $\sum_{i=1}^{k} \mu_i \nabla_\theta Q(\theta^{wc}, y_i) = 0$.

The necessary conditions for the solution of minimax problems can be derived from the classical theory of optimization in Banach spaces. Notice that Problem (1) can be expressed as $\min_{\theta \in \Theta, \, \rho \in \mathbb{R}} \{\rho : Q(\theta, y) \leq \rho, \forall y \in \mathbb{Y}\}$. The associate Lagrange function is defined as

$$
\begin{aligned}
L \; &= \; \rho + \int (Q(\theta^{wc}, y) - \rho) \, \mu(dy) + \sum_{j=1}^{p} \lambda_j h_j(\theta) \\
&= \; \int Q(\theta^{wc}, y) \, \mu(dy) + \sum_{j=1}^{p} \lambda_j h_j(\theta) + \rho \left(1 - \int \mu(dy)\right),
\end{aligned}
$$

where μ is a bounded Borel measure on $\mathbb{Y}$. Under an appropriate constraint qualification, the first order conditions of Problem (1) are:

$$
\int \nabla_\theta Q(\theta^{wc}, y) \, \mu(dy) + \sum_{j=1}^{p} \lambda_j \nabla_\theta h_j(\theta^{wc}) = 0, \quad 1 - \int \mu(dy) = 0, \quad (10)
$$

$$
\lambda_j h_j(\theta^{wc}) = 0, \; h_j(\theta^{wc}) \leq 0, \; \lambda_j \geq 0, \quad j = 1, ..., p,
$$

and $\int (Q(\theta^{wc}, y) - \rho) \, \mu(dy) = 0$, $Q(\theta^{wc}, y) - \rho \leq 0$. Therefore, as μ integrates one, $\rho = \int Q(\theta^{wc}, y) \, \mu(dy)$. Furthermore, $\sum_{j=1}^{p} \lambda_j h_j(\theta^{wc}) = 0$.

The Lagrange multiplier μ is a discrete measure with support in $\mathbb{Y}(\theta^{wc})$ and can be expressed as $\mu = \sum_{i=1}^{k} \mu_i \delta(y_i)$ for some $k \leq K + 1$ and $y_i \in \mathbb{Y}(\theta^{wc})$. This is because the set of measures μ satisfying the functional conditions (10) is convex, bounded and closed in the weak-* topology, and is therefore weakly-* compact. It follows from the Krein-Millman theorem that this set is equal to the convex hull of its extreme points. The extreme points can be shown to be discrete measures supported on $k \leq K + 1$ points because they satisfy $K + 1$ equations in (10), provided $\{\nabla_\theta h_j(\theta^{wc}) : h_j(\theta^{wc}) = 0\}$ are linearly independent vectors (see Shapiro (1994) and Shapiro (1998, pp. 112-113) for details). Alternative arguments based on the Caratéodory's Theorem can be found in Hager and Presler (1987). Because μ is a discrete probability measure, we can express (10) as,

$$
\sum_{i=1}^{k} \mu_i \nabla_\theta Q(\theta^{wc}, y_i) + \sum_{j=1}^{p} \lambda_j \nabla_\theta h_j(\theta^{wc}) = 0, \quad \sum_{i=1}^{k} \mu_i = 1.
$$

Any continuous minimax problem of the form (1) satisfying the assumptions in Theorem 9 can be written as:

$$
\min_{\rho, \theta \in \Theta} \{\rho : Q(\theta, y_i) \leq \rho, \quad i = 1, ..., k\}. \quad (11)
$$

The optima θ^{wc} and $\{y_i\}_{i=1}^{k} \subset \mathbb{Y}(\theta^{wc})$, and Lagrange's multipliers $\{\mu_i\}$, $\{\lambda_j\}$ of Problem (1) coincide with the optima and Lagrange's multipliers of Problem (11). This result will be applied to prove consistency of multipliers $\{\widehat{\mu}_i\}_{i=1}^{\widehat{k}}$ to $\{\mu_i\}_{i=1}^{k}$.

APPENDIX B: PROOFS

Proof of Theorem 1. The supremum and infimum on Euclidean spaces are measurable functions, as a consequence of the separability of Euclidean spaces. Then, the WC estimators can be chosen to be measurable (see e.g., Jennrich, 1969).

For any $\varepsilon > 0$, let define $B_\varepsilon = \{\theta \in \Theta, \|\theta - \theta^{wc}\| < \varepsilon\}$ and $B_\varepsilon^c = \Theta \backslash B_\varepsilon$. Then, as $\theta^{wc} \in B_\varepsilon$

$$
\begin{aligned}
\left\{\left\|\widehat{\theta}_T^{wc} - \theta^{wc}\right\| \geq \varepsilon\right\} &= \left\{\widehat{\theta}_T^{wc} \in B_\varepsilon^c\right\} \subset \left\{\inf_{\theta \in B_\varepsilon^c} \sup_{y \in \mathbb{Y}} Q_T(\theta, y) \leq \inf_{\theta \in B_\varepsilon} \sup_{y \in \mathbb{Y}} Q_T(\theta, y)\right\} \\
&\subset \left\{\inf_{\theta \in B_\varepsilon^c} \left(\sup_{y \in \mathbb{Y}} Q_T(\theta, y) - Q_T(\theta^{wc}, y^{wc})\right) \leq 0\right\} \\
&= \left\{\inf_{\theta \in B_\varepsilon^c} \sup_{y \in \mathbb{Y}} \left(K_{\theta^{wc}, y^{wc}}(\theta, y) - t_T(\theta, y)\right) \leq 0\right\} \\
&\subset \left\{\inf_{\theta \in B_\varepsilon^c} \sup_{y \in \mathbb{Y}} K_{\theta^{wc}, y^{wc}}(\theta, y) - \sup_{\theta \in \Theta} \inf_{y \in \mathbb{Y}} |t_T(\theta, y)| \leq 0\right\} \\
&\subset \left\{\sup_{\theta \in \Theta} \inf_{y \in \mathbb{Y}} |t_T(\theta, y)| > \delta\right\} \subset \left\{\sup_{\theta \in \Theta} \sup_{y \in \mathbb{Y}} |t_T(\theta, y)| > \delta\right\}
\end{aligned}
$$

but this sequence of events tends to zero almost surely (in probability) by assumption.

Proof of Proposition 2. We will prove that

$$
\sup_{y^{wc} \in \mathbb{Y}(\theta^{wc})} d\left(\mathbb{Y}_T\left(\widehat{\theta}_T^{wc}\right), y^{wc}\right) \to 0,
$$

for any $y^{wc} \in \mathbb{Y}(\theta^{wc})$ in Part 1. Part 2 proves the Proposition.

Part 1.

For any $y^{wc} \in \mathbb{Y}(\theta^{wc})$ and any $\varepsilon > 0$, we define the set $N_\varepsilon(y^{wc}) = \{y \in \mathbb{Y} : \|y - y^{wc}\| < \varepsilon\}$, and $N_\varepsilon(y^{wc})^c$ its complementary. Since $y^{wc} \in N_\varepsilon(y^{wc})$, then

$$
\begin{aligned}
&\bigcup_{y^{wc} \in \mathbb{Y}(\theta^{wc})} \left\{d\left(\mathbb{Y}_T\left(\widehat{\theta}_T^{wc}\right), y^{wc}\right) \geq \varepsilon\right\} = \bigcup_{y^{wc} \in \mathbb{Y}(\theta^{wc})} \left\{\mathbb{Y}_T\left(\widehat{\theta}_T^{wc}\right) \subset N_\varepsilon^c(y^{wc})\right\} \\
&\subset \bigcup_{y^{wc} \in \mathbb{Y}(\theta^{wc})} \left\{\sup_{y \in N_\varepsilon(y^{wc})^c} Q_T\left(\widehat{\theta}_T^{wc}, y\right) \geq \sup_{y \in N_\varepsilon(y^{wc})} Q_T\left(\widehat{\theta}_T^{wc}, y\right)\right\} \\
&\subset \bigcup_{y^{wc} \in \mathbb{Y}(\theta^{wc})} \left\{\sup_{y \in N_\varepsilon(y^{wc})^c} Q_T\left(\widehat{\theta}_T^{wc}, y\right) \geq Q_T\left(\widehat{\theta}_T^{wc}, y^{wc}\right)\right\} \\
&\subset \bigcup_{y^{wc} \in \mathbb{Y}(\theta^{wc})} \left\{\sup_{y \in N_\varepsilon(y^{wc})^c} \left(Q_T\left(\widehat{\theta}_T^{wc}, y\right) - Q_T\left(\widehat{\theta}_T^{wc}, y^{wc}\right)\right) \geq 0\right\} \\
&= \bigcup_{y^{wc} \in \mathbb{Y}(\theta^{wc})} \left\{\inf_{y \in N_\varepsilon(y^{wc})^c} \left(-Q_T\left(\widehat{\theta}_T^{wc}, y\right) + Q_T\left(\widehat{\theta}_T^{wc}, y^{wc}\right)\right) \leq 0\right\}.
\end{aligned}
$$

Therefore, under $A.1$, $A.2$, and $A.3$,

$$\bigcup_{y^{wc}\in\mathbb{Y}(\theta^{wc})}\left\{\inf_{y\in N_\varepsilon(y^{wc})^c}\left(Q_T\left(\widehat{\theta}_T^{wc},y^{wc}\right)-Q_T\left(\widehat{\theta}_T^{wc},y\right)\right)\leq 0\right\}$$

$$\subset\bigcup_{y^{wc}\in\mathbb{Y}(\theta^{wc})}\left\{\inf_{y\in N_\varepsilon(y^{wc})^c}\left(K_{\widehat{\theta}_T^{wc},y}\left(\widehat{\theta}_T^{wc},y^{wc}\right)-t_T\left(\widehat{\theta}_T^{wc},y^{wc}\right)\right)\leq 0\right\}$$

$$\subset\bigcup_{y^{wc}\in\mathbb{Y}(\theta^{wc})}\left\{\left|t_T\left(\widehat{\theta}_T^{wc},y^{wc}\right)\right|>\delta\right\}\subset\left\{\sup_{\theta\in\Theta}\sup_{y\in\mathbb{Y}}|t_T\left(\theta,y\right)|>\delta\right\}$$

and the result follows.

Part 2.

Note that $\left\{d_H\left(\mathbb{Y}_T\left(\widehat{\theta}_T^{wc}\right),\mathbb{Y}\left(\theta^{wc}\right)\right)>\varepsilon\right\}$ is equal to

$$\bigcup_{y^{wc}\in\mathbb{Y}(\theta^{wc})}\left\{d\left(\mathbb{Y}_T\left(\widehat{\theta}_T^{wc}\right),y^{wc}\right)>\varepsilon\right\}\bigcup_{\widehat{y}\in\mathbb{Y}_T\left(\widehat{\theta}_T^{wc}\right)}\left\{d\left(\widehat{y},\mathbb{Y}\left(\theta^{wc}\right)\right)>\varepsilon\right\}.$$

We have proved that the first set union in the right-hand side is included in

$$\left\{\sup_{\theta\in\Theta}\sup_{y\in\mathbb{Y}}|t_T\left(\theta,y\right)|>\delta\right\}.$$

Next we consider the second union of sets. Let $N_\varepsilon\left(\widehat{y}\right)=\{y\in\mathbb{Y}:\|y-\widehat{y}\|\leq\varepsilon\}$, and $N_\varepsilon\left(\widehat{y}\right)^c$ its complementary. Notice that

$$\bigcup_{\widehat{y}\in\mathbb{Y}_T\left(\widehat{\theta}_T^{wc}\right)}\left\{d\left(\widehat{y},\mathbb{Y}\left(\theta^{wc}\right)\right)>\varepsilon\right\}=\bigcup_{\widehat{y}\in\mathbb{Y}_T\left(\widehat{\theta}_T^{wc}\right)}\left\{\inf_{y^{wc}\in\mathbb{Y}(\theta^{wc})}\|\widehat{y}-y^{wc}\|>\varepsilon\right\}$$

$$\subset\bigcup_{\widehat{y}\in\mathbb{Y}_T\left(\widehat{\theta}_T^{wc}\right)}\left\{\sup_{y^{wc}\in\mathbb{Y}(\theta^{wc})\cap N_\varepsilon(\widehat{y})^c}Q_T\left(\widehat{\theta}_T^{wc},y^{wc}\right)\geq\sup_{y\in\mathbb{Y}(\theta^{wc})\cap N_\varepsilon(\widehat{y})}Q_T\left(\widehat{\theta}_T^{wc},y\right)\right\}$$

$$\subset\bigcup_{\widehat{y}\in\mathbb{Y}_T\left(\widehat{\theta}_T^{wc}\right)}\left\{\sup_{y^{wc}\in\mathbb{Y}(\theta^{wc})\cap N_\varepsilon(\widehat{y})^c}\left(Q_T\left(\widehat{\theta}_T^{wc},y^{wc}\right)-Q_T\left(\widehat{\theta}_T^{wc},\widehat{y}\right)\right)\geq 0\right\}$$

$$=\bigcup_{\widehat{y}\in\mathbb{Y}_T\left(\widehat{\theta}_T^{wc}\right)}\left\{\inf_{y^{wc}\in\mathbb{Y}(\theta^{wc})\cap N_\varepsilon(\widehat{y})^c}\left(Q_T\left(\widehat{\theta}_T^{wc},\widehat{y}\right)-Q_T\left(\widehat{\theta}_T^{wc},y^{wc}\right)\right)\leq 0\right\}$$

$$\subset\bigcup_{\widehat{y}\in\mathbb{Y}_T\left(\widehat{\theta}_T^{wc}\right)}\left\{\inf_{y^{wc}\in\mathbb{Y}(\theta^{wc})\cap N_\varepsilon(\widehat{y})^c}K_{\widehat{\theta}_T^{wc},y^{wc}}\left(\widehat{\theta}_T^{wc},\widehat{y}\right)-t_T\left(\widehat{\theta}_T^{wc},\widehat{y}\right)\leq 0\right\}$$

$$\subset\bigcup_{\widehat{y}\in\mathbb{Y}_T\left(\widehat{\theta}_T^{wc}\right)}\left\{\left|t_T\left(\widehat{\theta}_T^{wc},\widehat{y}\right)\right|>\delta\right\}\subset\left\{\sup_{\theta\in\Theta}\sup_{y\in\mathbb{Y}}|t_T\left(\theta,y\right)|>\delta\right\}$$

21

and the result follows.

Proof of Proposition 3. From Proposition 2, we can always take sets $\{\widehat{y}_i\}_{i=1}^{\widehat{k}} \subset \mathbb{Y}_T\left(\widehat{\theta}_T^{wc}\right)$, and $\{y_i\}_{i=1}^k \subset \mathbb{Y}(\theta^{wc})$ in such a way that $d_H\left(\{\widehat{y}_i\}_{i=1}^{\widehat{k}}, \{y_i\}_{i=1}^k\right) \to 0$, and $\widehat{k} \to k$, almost surely (in probability), without loss of generality. As $\{\mu_i\}_{i=1}^k$ are the Lagrange multipliers associated with the problem

$$\min_{\rho} \{\rho : Q\left(\theta^{wc}, y_i\right) \leq \rho, \quad i = 1, ..., k\}$$

and $\{\widehat{\mu}_i\}$ are Lagrange multipliers associated with the problem

$$\min_{\rho} \left\{\rho : Q_T\left(\widehat{\theta}_T^{wc}, \widehat{y}_i\right) \leq \rho, \quad i = 1, ..., \widehat{k}\right\},$$

it is sufficient to check that the Lagrange functions associated with these two problems,

$$
\begin{aligned}
L\left(\rho, \mu\right) &= \rho - \sum_{i=1}^k \mu_i \left(Q\left(\theta^{wc}, y_i\right) - \rho\right) \\
L_T\left(\rho, \mu\right) &= \rho - \sum_{i=1}^{\widehat{k}} \mu_i \left(Q_T\left(\widehat{\theta}_T^{wc}, \widehat{y}_i\right) - \rho\right),
\end{aligned}
$$

converge uniformly. Since that $\widehat{k} \to k$, and $\sum_{i=1}^k \mu_i = 1$, with nonnegative μ_i, the Kolmogorov distance between Lagrange functions satisfies,

$$
\begin{aligned}
\sup_{\rho,\mu}\left|L_T\left(\rho, \mu\right) - L\left(\rho, \mu\right)\right| &= \sup_{\rho,\mu}\left|\sum_{i=1}^k \mu_i\left(Q_T\left(\widehat{\theta}_T^{wc}, \widehat{y}_i\right) - Q\left(\theta^{wc}, y_i\right)\right)\right| + o\left(1\right) \\
&\leq \max_{y \in \mathbb{Y}}\left|Q_T\left(\widehat{\theta}_T^{wc}, y\right) - Q\left(\theta^{wc}, y\right)\right| + o\left(1\right),
\end{aligned}
$$

where the $o\left(1\right)$ term is uniform in ρ, μ. Next,

$$\max_{y \in \mathbb{Y}}\left|Q_T\left(\widehat{\theta}_T^{wc}, y\right) - Q\left(\theta^{wc}, y\right)\right| \to 0,$$

when $\max_{\theta \in \Theta, y \in \mathbb{Y}}\left|Q_T\left(\theta, y\right) - Q\left(\theta, y\right)\right| \to 0$, and $\widehat{\theta}_T \to \theta^{wc}$ almost surely (in probability). The result follows from a standard application of the Consistency Theorem for extreme estimators on a compact domain (the positive simplex in $\mathbb{R}^k$, and any compact interval containing the optima $\rho^* = Q\left(\theta^{wc}, y_i\right)$ for all i).

Proof of Theorem 4. Given an arbitrary vector δ such that $\delta'\delta = 1$, let

$$d_T = \delta'\left(B^{-1}AB^{-1}\right)^{-1/2}\sqrt{T}\left(\widehat{\theta}_T^{wc} - \theta^{wc}\right).$$

If $\theta^{wc} \in int\,\{\Theta\}$ solves Problem (1) and $\widehat{\theta}_T^{wc}$ is a consistent estimator, then $\Pr\left(\widehat{\theta}_T^{wc} \notin int\,\{\Theta\}\right) \to$ 0. Therefore,

$$
\begin{aligned}
\Pr\left(d_T \le x\right) &= \Pr\left(d_T \le x \mid \widehat{\theta}_T^{wc} \in int\,\{\Theta\}\right) + \\
&+ \left[\Pr\left(d_T \le x \mid \widehat{\theta}_T^{wc} \notin int\,\{\Theta\}\right) - \Pr\left(d_T \le x \mid \widehat{\theta}_T^{wc} \in int\,\{\Theta\}\right)\right] \Pr\left(\widehat{\theta}_T^{wc} \notin int\,\{\Theta\}\right) \\
&= \Pr\left(d_T \le x \mid \widehat{\theta}_T^{wc} \in int\,\{\Theta\}\right) + o\,(1)
\end{aligned}
$$

uniformly in x.

By Theorem 9, there exists a positive integer $1 \le \widehat{k} \le K+1$, vectors $\widehat{y}_i \in \mathbb{Y}\left(\widehat{\theta}_T^{wc}\right)$ and multipliers $\widehat{\mu}_i \ge 0$ for $i = 1, ..., \widehat{k}$ with $\sum_{i=1}^{k} \widehat{\mu}_i = 1$, such that

$$
\sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \nabla_\theta Q_T \left(\widehat{\theta}_T^{wc}, \widehat{y}_i\right) = 0.
$$

Applying the mean value theorem,

$$
\begin{aligned}
0 &= \sqrt{T} \sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \frac{\partial Q_T\left(\widehat{\theta}_T^{wc}, \widehat{y}_i\right)}{\partial \theta} \\
&= \sqrt{T} \sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \frac{\partial Q_T\left(\theta^{wc}, \widehat{y}_i\right)}{\partial \theta} + \sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \frac{\partial^2 Q_T\left(\tilde{\theta}_T, \widehat{y}_i\right)}{\partial \theta \partial \theta'} \sqrt{T}\left(\widehat{\theta}_T^{wc} - \theta^{wc}\right),
\end{aligned}
$$

where $\left\|\tilde{\theta}_T - \theta^{wc}\right\| \le \left\|\widehat{\theta}_T - \theta^{wc}\right\|$. Under $B.4.$, it follows that,

$$
0 = \sqrt{T} \sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \frac{\partial Q_T\left(\theta^{wc}, \widehat{y}_i\right)}{\partial \theta} + \left[B + o_p\,(1)\right]\sqrt{T}\left(\widehat{\theta}_T^{wc} - \theta^{wc}\right).
$$

Using Conditions $B.2.$ and $B.3.$,

$$
\sqrt{T} \sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \frac{\partial Q_T\left(\theta^{wc}, \widehat{y}_i\right)}{\partial \theta} \to_d N_p\,(0, A).
$$

Thus,

$$
B^{-1}\left[B + o_p\,(1)\right]\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) = -B^{-1}\left\{\sqrt{T} \sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \frac{\partial Q_T\left(\theta^{wc}, \widehat{y}_i\right)}{\partial \theta}\right\} \to_d N\left(0, B^{-1}AB^{-1}\right)
$$

and the result follows.

Proof of Proposition 5. By Condition $C.2$,

$$E\left[\sup_{\|\theta-\theta^{wc}\|\leq\delta}\left\|\sum_{i=1}^{\widehat{k}}\widehat{\mu}_i\left(\frac{\partial^2 Q_T\left(\theta,\widehat{y}_i\right)}{\partial\theta\partial\theta'}-\frac{\partial^2 Q_T\left(\theta^{wc},\widehat{y}_i\right)}{\partial\theta\partial\theta'}\right)\right\|\right]$$

$$\leq\quad E\left[\sup_{\|\theta-\theta^{wc}\|\leq\delta}\sup_{y\in\mathbb{Y}}\left\|\frac{\partial^2 Q_T\left(\theta,y\right)}{\partial\theta\partial\theta'}-\frac{\partial^2 Q_T\left(\theta^{wc},y\right)}{\partial\theta\partial\theta'}\right\|\sum_{i=1}^{\widehat{k}}\widehat{\mu}_i\right]_\delta\underset{\delta\downarrow 0}{\rightarrow}0;$$

as $\sum_{i=1}^{\widehat{k}}\widehat{\mu}_i=1$.

Next, we use that $\forall\varepsilon>0$,

$$\Pr\left(\left\|\sum_{i=1}^{\widehat{k}}\widehat{\mu}_i\frac{\partial^2 Q_T\left(\tilde{\theta},\widehat{y}_i\right)}{\partial\theta\partial\theta'}-B\right\|>\varepsilon\right)\leq\Pr\left(\left\|\sum_{i=1}^{\widehat{k}}\widehat{\mu}_i\frac{\partial^2 Q_T\left(\theta^{wc},\widehat{y}_i\right)}{\partial\theta\partial\theta'}-B\right\|>\frac{\varepsilon}{2}\right)+$$

$$\Pr\left(\left\|\sum_{i=1}^{\widehat{k}}\widehat{\mu}_i\left(\frac{\partial^2 Q_T\left(\tilde{\theta},\widehat{y}_i\right)}{\partial\theta\partial\theta'}-\frac{\partial^2 Q_T\left(\theta^{wc},\widehat{y}_i\right)}{\partial\theta\partial\theta'}\right)\right\|>\frac{\varepsilon}{2}\right).$$

The first term tends to zero by Condition $C.1$. The second term tends to zero by Condition $C.2$ and Markov's inequality, since $\left\|\tilde{\theta}_T-\theta^{wc}\right\|\rightarrow_p 0$, so we can build a sequence $\delta_T\rightarrow 0$ such that $\left\|\tilde{\theta}_T-\theta^{wc}\right\|\leq\delta_T$ except for sets of probability tending to zero.

Proof of Theorem 6. The proof is analogous to that of Theorem 4. Assuming $B.1'$ and h is continuous, it is satisfied that $h_j\left(\widehat{\theta}_T^{wc}\right)=0$ for $j=1,...,p$, except for a set of probability tending to zero. Applying the mean value theorem to the first order necessary conditions,

$$\sum_{i=1}^{\widehat{k}}\widehat{\mu}_i\frac{\partial}{\partial\theta}Q_T\left(\widehat{\theta}^{wc},\widehat{y}_i\right)+\sum_{j=1}^{p}\widehat{\lambda}_j\frac{\partial}{\partial\theta}h_j\left(\widehat{\theta}^{wc}\right)\quad=\quad 0,$$

$$h_j\left(\widehat{\theta}^{wc}\right)\quad=\quad 0,\qquad j=1,...,p$$

we obtain,

$$
\begin{pmatrix} \sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \frac{\partial}{\partial \theta} Q_T\left(\widetilde{\theta}, \widehat{y}_i\right) & \frac{\partial}{\partial \theta} H_p\left(\widetilde{\theta}\right)' \\ \frac{\partial}{\partial \theta} H_p\left(\widetilde{\theta}\right) & 0 \end{pmatrix} \sqrt{T} \begin{pmatrix} \widehat{\theta}_T^{wc} - \theta^{wc} \\ \widehat{\lambda}^{wc} \end{pmatrix}
$$

$$
= \begin{pmatrix} \sqrt{T} \sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \frac{\partial}{\partial \theta} Q_T\left(\theta^{wc}, \widehat{y}_i\right) \\ 0 \end{pmatrix}
$$

where $H_p\left(\theta\right) = \left(h_1\left(\theta\right), ..., h_p\left(\theta\right)\right)'$. The asymptotic normality follows analogously to Theorem 4, with covariance matrix V equal to

$$
V = \begin{pmatrix} B & H' \\ H & 0 \end{pmatrix}^{-1} \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} B & H' \\ H & 0 \end{pmatrix}^{-1},
$$

where $H = \partial H_p\left(\theta^{wc}\right)/\partial \theta$.

Proof of Theorem 7. The proof is similar to that of Theorem 4. Under conditions $D.1$, $D.2$, $D.3$ and $D.4$

$$
\sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \left(T^{-1} \sum_{t=1}^{T} \frac{\partial}{\partial \theta} g\left(X_t, \widehat{y}_i, \widehat{\theta}_T^{wc}\right)\right) W_T \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g\left(X_t, \widehat{y}_i, \theta^{wc}\right)\right) \to_d N\left(0, D\right)
$$

applying the delta method, and for any sequence $\widetilde{\theta}_T \to_p \theta^{wc}$,

$$
\sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \left(T^{-1} \sum_{t=1}^{T} \frac{\partial}{\partial \theta} g\left(X_t, \widehat{y}_i, \widetilde{\theta}_T\right)\right) W_T \left(T^{-1} \sum_{t=1}^{T} \frac{\partial}{\partial \theta} g\left(X_t, \widehat{y}_i, \widetilde{\theta}_T\right)\right) \to_p E.
$$

When $\widehat{\theta}_T \in int\left\{\Theta\right\}$,

$$
\begin{aligned}
0 &= 2\sqrt{T} \sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \frac{\partial Q_T\left(\widehat{\theta}_T^{wc}, \widehat{y}_i\right)}{\partial \theta} \\
&= \sqrt{T} \sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \theta} g\left(X_t, \widehat{y}_i, \widehat{\theta}_T\right)\right) W_T \left(\frac{1}{T} \sum_{t=1}^{T} g\left(X_t, \widehat{y}_i, \widehat{\theta}_T\right)\right).
\end{aligned}
$$

Applying the mean value theorem,

$$
g\left(X_t, \widehat{y}_i, \widehat{\theta}_T\right) = g\left(X, \widehat{y}_i, \theta^{wc}\right) + \frac{\partial}{\partial \theta} g\left(X_t, \widehat{y}_i, \widetilde{\theta}_T\right)\left(\widehat{\theta}_T - \theta^{wc}\right),
$$

25

with $\left\|\tilde{\theta}_T - \theta^{wc}\right\| \leq \left\|\widehat{\theta}_T - \theta^{wc}\right\|$, and therefore

$$
\begin{aligned}
0 &= \sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \theta} g \left(X_t, \widehat{y}_i, \widehat{\theta}_T \right) \right) W_T \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g \left(X_t, \widehat{y}_i, \theta^{wc} \right) \right) \\
&+ \sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \theta} g \left(X_t, \widehat{y}_i, \widehat{\theta}_T \right) \right) W_T \left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \theta} g \left(X_t, \widehat{y}_i, \tilde{\theta}_T \right) \right) \sqrt{T} \left(\widehat{\theta}_T^{wc} - \theta^{wc} \right).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
& E^{-1} \left[E + o_p \left(1 \right) \right] \sqrt{T} \left(\widehat{\theta}_T^{wc} - \theta^{wc} \right) \\
&= -E^{-1} \sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \theta} g \left(X_t, \widehat{y}_i, \widehat{\theta}_T \right) \right) W_T \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g \left(X_t, \widehat{y}_i, \theta^{wc} \right) \right) \\
&\rightarrow {}_d N \left(0, E^{-1} D E^{-1} \right).
\end{aligned}
$$

REFERENCES

Aigner, D. J. C. (1974) An Appropriate Econometric Framework for Estimating a Labour-Supply Function from the SEO File. *International Economic Review* 15, 59-68.

Aigner, D. J. C., C. Hsiao, A. Kapteeyn & T. Wansbeek (1984) Latent Variable Models in Econometrics. In Z. Griliches & M. D. Intriligator (eds.), *Handbook of Econometrics*, Vol. II, pp. 1323-1393. Amsterdam: North Holland.

Amemiya, Y. (1985a) Instrumental Variable Estimator for the Nonlinear Errors-in-Variables Model. *Journal of Econometrics* 28, 273-289.

Amemiya, Y. (1985b) *Advanced Econometrics.* Cambridge: Harvard University Press.

Bector, C. R. & B. L. Bhatia (1985) Sufficient Optimality Conditions and Duality for a Minimax Problem. *Utilitas Mathematica* 27, 229-247.

Berge, C. (1963) Topological Spaces. New York: Macmillan.

Bickel, P. J. & M. J. Wichura (1971) Convergence Criteria for Multiparameter Processes and Some Applications. *The Annals of Mathematical Statistics* 42, 1656-1670.

Billingsley, P. (1968) Convergence of Probability Measures. New York: John Wiley and Sons. 2nd. edition in 1999.

Bierens, H. J. (1981) Robust Methods and Asymptotic Theory in Nonlinear Econometrics. Lecture Notes in Economics and Mathematical Systems, Vol. 192. Berlin: Springer Verlag.

Bierens, H. J. (1994) Topics in Advanced Econometrics: Estimation. Testing and Specification of Cross-section and Time Series Models. Cambridge: Cambridge University Press.

Bram, J. (1966) The Lagrange Multiplier Theorem for Max-Min with Several Constraints. *SIAM Journal of Applied Mathematics* 14 (4), 665-667.

Bowden, R. J. & D. Turkington (1984) Instrumental Variables, Econometric Society Monographs in Quantitative Economics. Cambridge: Cambridge University Press.

Chamberlain, G. (1977) Education, Income & Ability Revisited. *Journal of Econometrics* 5, 241-257.

Chamberlain, G. & Z. Griliches (1975) Unobservables with a Variance Components Structure: Ability, Schooling and the Economic Success of Brothers. *International Economic Review* 16, 422-449.

Danskin, J. M. (1967) The Theory of Max-Min. Berlin: Springer.

Davidson, J. (1994) Stochastic Limit Theory, Advanced Text in Econometrics. Oxford: Oxford University Press.

Dempster, A., N. Laird & D. Rubin (1977) Maximum Likelihood from Incomplete Data via EM Algorithm. *Journal of the Royal Statistical Society* (Ser. B) 39, 1-38.

Dem'yanov, V. F. & V. N. Malozemov (1972) Vvedenie Minimaks. Moscow: Nauka. In English, 1974, Introduction to Minimax. New York: Wiley.

Devroye, L. (1987) A Course in Density Estimation. Progress in Probability and Statistics. Boston: Birkhäuser.

Dudley, R. M. (1999) Uniform Central Limit Theorems. Cambridge: Cambridge University Press.

Fuller, W. A. (1987) Measurement Error Models. New York: J. Wiley & Sons.

Gabler, S. (1990) Minimax Solutions in Sampling from Finite Populations. Lecture Notes in Statistics, 64. New York: Springer Verlag.

Goldberger, A. S. (1972) Maximum Likelihood Estimation of Regressions Containing Unobservable Independent Variables. *International Economic Review* 13, 1-15.

Goldberger, A. S. (1974) Unobservable Variables in Econometrics. In P. Zarembka (ed.), *Frontiers in Econometrics.* New York: Academic Press.

Griliches, Z. (1974) Errors in Variables and Other Unobservables. *Econometrica* 42, 971-998.

Griliches, Z. (1977) Estimating the Returns to Schooling: Some Econometric Problems. *Econometrica* 45, 1-22.

Griliches, Z. & W. M. Mason (1972) Education and Ability. *Journal of Political Economy* 80, 74-103.

Hager, W. W. & D. L. Presler (1987) Dual Techniques for Minimax. S*IAM Journal. of Control and Optimization* 25, 660-685.

Hall, A. (1993) Some Aspects of Generalized Method of Moments Estimation. In: G. S. Maddala & C. R. Rao (eds.), *Handbook of Statistics*, Vol. 11, pp. 393–417. Elsevier Science Pub. Co.

Hansen, L. P. (1982) Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* 50, 1029-1054.

Hansen, L. P. (1985) A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators. *Journal of Econometrics* 30, 203-238.

Hansen, L. P. & K. J. Singleton (1982) Generalized Instrumental Variables Estimation of Nonlinear Rational Expectation Models. *Econometrica* 50, 1269-1286.

Hansen, L. P. & T. J. Sargent (2000). Wanting Robustness in Macroeconomics. Stanford University, mimeo.

Hausman, J. A., J. Abrevaya & F. M. Scott-Morton (1998) Misclassification of the Dependent Variable in a Discrete Response Setting. *Journal of Econometrics* 87, 237-269.

Hausman, J. A., W. K. Newey, H. Ichimura & J. L. Powell (1991) Identification and Estimation of Polynomial Errors-in-Variables Models. *Journal of Econometrics* 50, 273-295.

Hausman, J. A., W. K. Newey & J. L. Powell (1995) Nonlinear Errors in Variables Estimation of Some Engle Curves. *Journal of Econometrics* 65, 205-233.

Hettich, R. & K. O. Kortanek (1993) Semi-infinite Programming: Theory, Methods and Applications. *SIAM Review* 35, 380-429.

Hsiao, C. & B. Sun (1999) Modelling Survey Response Bias -with and Analysis of the Demand for an Advanced Electronic Device. *Journal of Econometrics* 89, 15-39.

Huber, P.J. (1964) Robust estimation of location parameter. *Annals of Mathematical Statistics* 37, 73-101.

Huber, P.J. (1967) The behavior of Maximum Likelihood Estimates under Non-Standard Conditions. In: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press.

Huber, P.J. (1994) Robust Statistical Procedures (Second ed.), CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 68. Philadelphia: SIAM.

Hum, D. & W. Simpson (1994) Labour Supply Estimation and Public Policy. *Journal of Economic Surveys* 8, 57-81.

Jennrich, R. I. (1969) Asymptotic Properties of Non Linear Least Squares Estimators. *Annals of Mathematical Statistics* 40, 633-643.

Jöreskog, K. G. (1973) A General Method for Estimating a Linear Structural Equation System. In: Goldberger, A. S. & O. D. Duncan (eds.), *Structural Equation Models in Mathematical Psychology*, Vol. 2, pp. 85-112. New York: Seminar.

Jöreskog, K. G. (1978) Statistical Analysis of Covariance and Correlation Matrices. *Psychometrica* 43, 443-447.

Lehmann, E. L. (1983) Theory of Point Estimation. New York: J. Wiley & Sons.

Lehmann, E. L. (1986) Testing Statistical Hypotheses, 2nd edition. New York: J. Wiley & Sons.

Li, T., P. K. Trivedi & J. Guo (2003) Modelling Response Bias in Count: a Structural Approach with an Application to National Crime Victimization Survey Data. *Sociological Methods and Research* 31, 514-544.

Li, T. & C. Hsiao (2004) Robust Estimation of Generalized Linear Models with Measurement Errors. *Journal of Econometrics* 118, 51-65.

Lucas, R. E. (1976) Econometric Policy Evaluation: a Critique. In: K. Bruckner & A. Meltzer (eds.), *The Phillips Curve and Labour Markets*, Vol. 1 of Carnegie-Rochester Conferences on Public Policy, pp. 19-46. Amsterdam: North Holland.

Manski, C. (1994) Analog Estimation of Econometric Models. In R. F. Engle & D. L. McFadden (eds.), *Handbook of Econometrics*, Vol. IV. Amsterdam: North Holland.

Newey, W. K. (1993) Efficient Estimation of Models with Conditional Moment Restrictions. In: G. S. Maddala & C. R. Rao, (eds.), *Handbook of Statistics*, Vol. 11, pp. 419-453. Elsevier Science Pub Co.

Newey, W. K. & D. McFadden (1994) Large Sample Estimation and Hypothesis Testing. In: R. F. Engle & D. L. McFadden (eds.), *Handbook of Econometrics*, Vol. IV, pp. 419-453. Amsterdam: North Holland.

Ogaki, M. (1993) Generalized Method of Moments: Econometric Applications. In G. S. Maddala & C. R. Rao (eds.), *Handbook of Statistics*, Vol. 11, pp. 455-488. Elsevier Science Pub Co.

Pollard, D. (1989) Asymptotics via Empirical Processes. *Statistical Science* 4, 341-366.

Pollard, D. (1990) Empirical Processees: Theory and Applications, NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 2. Hayward, CA: Institute of Mathematical Statistics.

Pötscher, B. M. & I. R. Prucha (1991a) Basic Structure of the Asymptotic Theory in Dynamic Nonlinear Econometric Models, Part I: Consistency and Approximation Concepts. *Econometric Reviews* 10, 125-216.

Pötscher, B. M. & I. R. Prucha (1991b) Basic Structure of the Asymptotic Theory in Dynamic Nonlinear Econometric Models, Part II: Asymptotic Normality. *Econometric Reviews* 10, 253-325.

Pötscher, B. M. & I. R. Prucha (1997) Dynamic Nonlinear Econometric Models. New York: Springer Verlag.

Reemtsen, R. & S. Görner (1998) Numerical Methods for Semi-Infinite Programming: a Survey. In R. Reemtensen & J. J. Rückmann (eds.), *Semi-infinite Programming,* pp. 195-275. Dordretch, The Netherlands: Kluwer Academic Press.

Robinson, P. M. (1974) Identification, Estimation and Large-Sample Theory for Regressions Containing Unobservable Variables. *International Economic Review* 15, 680-692.

Rockafellar, R. T. (1970) Convex Analysis. Princeton: Princeton University Press.

Rosen, J. B. (1965) Existence and Uniqueness of Equilibrium Points for Concave n-Person Games. *Econometrica* 33, 520-534.

Rustem, B. & M. A. Howe (2002) Algorithms for Worst-Case Design and Applications to Risk Management. Princeton and Oxford: Princeton University Press.

Sargan, J. D. (1958) The Estimation of Economic Relationship Using Instrumental Variables. *Econometrica* 26, 393-415.

Schmitendorf, W. E. (1977) Necessary Conditions and Sufficient Conditions for Static Min-Max Problems. *Journal of Mathematical Analysis and Applications* 57, 683-693.

Shapiro, A. (1994) On Lipschitzian stability of optimal solutions of parametrized semiinfinite programs. *Mathematics of Operation Research* 19, 743-752.

Shapiro, A. (1988) First and Second Order Optimality Conditions and Perturbation Analysis of Semi-Infinite Programming Problems. In: R. Reemtensen & J. J. Rückmann (eds.), *Semi-infinite Programming,* pp. 103-133. Dordretch, The Netherlands: Kluwer Academic Press.

Shimizu, K. & E. Aiyoshi (1980) Necessary Conditions for Min-Max Problems and Algorithms by a Relaxation Procedure. *IEE Transactions on Automatic Control* 25, 62-66.

Sundaram, R. K. (1996) A First Course in Optimization Theory. New York: Cambridge University Press.

van der Vaart, A. W. (1998) Asymptotic Statistics. Cambridge: Cambridge University Press.

van der Vaart, A. W. & J. A. Wellner (1996) Weak Convergence and Empirical Processes with Applications to Statistics. New York: Springer Verlag.

van Geer, S. (2000) Empirical Processes in M-Estimation. Cambridge: Cambridge University Press..

Wei, G. C. G. & M. A. Tanner (1990) A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms. *Journal of the American Statistical Association* 88, 699-704.

Wichura, M. J. (1969) Inequalities with Applications to the Weak Convergence of Random Processes with Multidimensional Time Parameters. *The Annals of Mathematical Statistics* 40, 681-687.

Wooldridge, J. M. (1994) Estimation and Inference for Dependent Processes. In: R. F. Engle & D. L. McFadden (eds.), Handbook of Econometrics, Vol. IV, pp. 2639-2738. Amsterdam: North Holland.

Žaković, S. & B. Rustem (2003) Semi-infinite Programming and Applications to Minimax Problems. *Annals of Operations Research* 124, 81-110.

Zellner, A. (1970) Estimation of Regression Relationships Containing Unobservable Variables. *International Economic Review* 11, 441-654.