

# Worst-case estimation for econometric models with unobservable components

Mercedes Esteban-Bravo, Jose M. Vidal-Sanz\*

*Department of Business Administration, Universidad Carlos III de Madrid, C/ Madrid 126, Getafe, Madrid, Spain*

---

## Abstract

A worst-case estimator for econometric models containing unobservable components, based on minimax principles for optimal selection of parameters, is proposed. Worst-case estimators are robust against the adverse effects of unobservables. Computing worst-case estimators involves solving a minimax continuous problem, which is quite a challenging task. Large sample theory is considered, and a Monte Carlo study of finite-sample properties is conducted. A financial application is considered.

*Keywords:* Worst-case decision; Robust modelling; Minimax optimization

---

## 1. Introduction

The use of incomplete data has long been an issue in applied economics, and still there is no consensus about which inference methodology should be considered to calibrate the parameters of economic models. 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 identification 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); and then, the efficiency of the estimation becomes the central matter of concern to statisticians, as Horowitz and Manski (2006) point out. However, the consistency of these estimators is conditioned to the validity of the postulated hypotheses. Nonobservability renders the diagnosis of these hypotheses difficult to implement, and this limitation sometimes leads to the “illegitimate” rejection of economic theories.

We can certainly estimate pseudo parameters in models under wrong identification assumptions, (see, e.g., White, 1994), but the risk associated to the use of these estimations in economic models is unlimited. For instance, Aigner (1974) uses the 1967 Survey of Economic Opportunities to estimate the labor-supply function as an errors in variables model. Hum and Simpson (1994) suggest that a bias in labor-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.

---

\* Corresponding author. Tel.: +34 91 6248642; fax: +34 91 6249607.

*E-mail addresses:* mercedes.esteban@uc3m.es (M. Esteban-Bravo), jvidal@emp.uc3m.es (J.M. Vidal-Sanz).

In view of frequently expressed scepticism over untestable identification hypotheses in models containing unobservables, recent research has focused on conservative inferences enhancing the computation of finite sample bounds for the unidentified estimators (see, e.g., Horowitz and Manski, 1998, 2000, 2006; Horowitz et al., 2003; Imbens and Manski, 2004; Chernozhukov et al., 2004; Honoré and Tamer, 2005; see also Fishman and Rubin, 1998). However, in the case of wide bounds there is little information about the parameter (Horowitz et al., 2003, p. 421); and again, the parameter can be only determined invoking untestable assumptions.

The aim of this paper is to present a robust approach based on minimax principles for optimal selection and estimation of parameters in models with unobservables. In other words, we propose a worst-case (WC) strategy to seek optimal estimators—that we will call “worst-case” estimators—in the WC value of the unobservable components. The *WC estimation procedure guarantees best upper-bound loss in view of the WC 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. Under appropriate conditions, we prove consistency and asymptotic normality of WC estimators. From a computational point of view, computing WC estimators is quite a challenging task, as it involves 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). We use the global optimization algorithm considered by Žaković and Rustem (2003). This approach first specifies an equivalent semi-infinite programming problem to the original problem and then solves the semi-infinite programming problem by a global optimization approach.

WC techniques have also been appreciated in different economic contexts. Minimax principles have been applied in game theory in the study of decision making in  $n$ -person conflicts (see e.g., Rosen, 1965). In a WC 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 WC strategy seeks optimal responses in the WC value of uncertainty. WC approach has also been appreciated in finance, with applications in portfolio management, see e.g. Balbás and Ibañez (2002), Rustem and Howe (1997, 2002). The use of minimax approaches for designing robust economic policies has been considered by Hansen and Sargent (2001), linking max–min utility theory and robust control theory. The robust design of monetary policies based on minimax optimization have been considered by Rustem et al. (2005, 2002), Tetlow and von zur Muehlen (2001) and the unpublished monograph of Hansen and Sargent (2005), among others.

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), optimal sampling designs from finite populations (see e.g., Gabler, 1990), robust estimation of wavelet regression models with unknown disturbance components (see e.g., Tian and Herzberg, 2006), computation of upper and lower bounds for variances of discrete variables (Fishman and Rubin, 1998) and robust Bayesian analysis minimizing the expected loss for the WC prior (Noubiap and Seidel, 2001). Huber (1964) introduces a groundbreaking robust method of estimating location parameters for contaminated normal distributions, minimizing the maximal (WC) asymptotic variance that can happen over a neighborhood of the specified model (see also Huber, 1994, Chapter IV).

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 WC 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. In the concluding section we summarize the findings and discuss how some regularity assumptions can be circumvented. Proofs are placed in the appendix.

## 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. Let  $\Theta \subset \mathbb{R}^K$  be a compact set of parameters and  $(X, Y)$  be a

vector of random variables. Parameters  $\theta^0$  are defined as the minimizers of a loss function  $Q(\theta) = E[g(X, Y, \theta)]$  on  $\Theta$ , where  $g$  is a continuous function. Following the *analogy principle*, given a sample  $\{X_t, Y_t\}_{t=1}^T$  identically distributed as  $(X, Y)$ , parameters  $\theta^0$  can be consistently estimated by minimizing  $Q_T(\theta) = T^{-1} \sum_{t=1}^T g(X_t, Y_t, \theta)$  on  $\Theta$ . The minimizer  $\hat{\theta}_T$  is known as M-estimator.

As discussed in the Introduction, many theoretical models  $g(X, Y, \theta)$  are given by unknown factors  $Y$  that cannot be easily observed, even though we roughly know their variation range  $\mathbb{Y} \subset \mathbb{R}^S$ . Examples are:

- (1) Nuisance parameters  $Y \in \mathbb{Y}$  determined exogenously to the model, and whose values are not available or out-dated. For example, when  $Y$  is a parameter which is expected to change due to some exogenous structural change (that cannot be completely anticipated), and the researcher seeks a robust model against this unknown.
- (2) Unobserved random variables  $Y_t \in \mathbb{Y}$  for which we dare not make any distributional assumption, nor postulate any proxy and/or instruments.

In these cases the model is unidentified, and we propose a robust estimation approach based on minimax principles for optimal selection of parameters associated to the unobservable component (i.e., we minimize the WC value of the loss function with respect to the unobservables).

Let  $Q(\theta, y) = E[g(X, y, \theta)]$  denote the WC loss function, where  $\theta \in \Theta$  and  $y$  is a vector of unobservable components defined on  $\mathbb{Y} \subset \mathbb{R}^S$ . The WC strategy considers parameters  $\theta^{\text{wc}}$  that solves the problem

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

These parameters  $\theta^{\text{wc}}$  are those that best fit the available data in view of the unobservable component  $Y$ . The WC strategy safeguards against the WC outcomes of the unobservable element  $Y$  and makes no assumptions about the statistical nature of  $Y$ . Given a sample  $\{X_t\}_{t=1}^T$  identically distributed as  $X$ , and the sample analog of  $Q(\theta, y)$ ,

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

the WC estimator  $\hat{\theta}_T^{\text{wc}}$  of  $\theta^{\text{wc}}$  is defined as the solution to

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

The WC 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.

Here we discuss the reasons why risk averse modellers should consider WC estimators and balance the resulting estimates to determine a robust model. Although the true loss  $Q(\cdot)$  is unidentified when  $Y$  is unobserved, WC parameters guarantee that the loss  $Q(\theta^{\text{wc}})$  is bounded, as we show next. If  $Y$  are deterministic nuisance parameters (a degenerate random variable), it is trivially satisfied that

$$Q(\theta^{\text{wc}}) = E[g(X, Y, \theta^{\text{wc}})] \leq \max_{y \in \mathbb{Y}} E[g(X, y, \theta^{\text{wc}})] = \max_{y \in \mathbb{Y}} Q(\theta^{\text{wc}}, y).$$

Therefore, the WC parameters guarantee an improvement of the true loss function no matter which is the unknown  $Y$ , the true loss value associated with  $\theta^{\text{wc}}$  is upper bounded, and the bound can be estimated by  $\max_{y \in \mathbb{Y}} Q_T(\hat{\theta}_T^{\text{wc}}, y)$ . If  $Y$  is stochastic, the loss is also bounded under proper conditions. Let  $F_0(x, y)$  be the probability distribution of  $(X, Y)$ . We introduce the following assumptions:

- H.0.  $F_0(x, y)$  is absolutely continuous, and the conditional density satisfies  $f_0(y|x) \leq k$  almost surely (and let us define  $c = k \cdot \mu(\mathbb{Y})$ , where  $\mu$  is the Lebesgue measure).
- H.0'. There exists a  $c > 0$  such that  $\sup_{x, y} \{F_0(x, y)/F_0(y)F_0(x) : F_0(y) > 0, F_0(x) > 0\} < c$ .

Condition H.0' is satisfied if  $X, Y$  are independent and then  $c = 1$ .

**Lemma 1.** Assume that  $F_0(x, y)$  satisfies H.0 or H.0', and  $g$  is nonnegative. Then

$$Q(\theta^{\text{wc}}) \leq c \max_{y \in \mathbb{Y}} Q(\theta^{\text{wc}}, y).$$

The proof is left to the appendix.

As discussed in the Introduction, a large body of econometrics and statistics literature has addressed the estimation of models containing unobservable random variables, assuming a probability distribution or a reduced form model for the unobservable. However, economists rarely have information on the probability law of the unobservables, and the choice of their distribution is typically a matter of convenience rather than an expression of actual knowledge. If this assumption is invalid, estimators are usually inconsistent (i.e. estimations converge in probability to a parameter  $\theta^1 \neq \theta^0$ ). Furthermore, the loss of the estimated parameter  $Q(\theta^1)$  can be arbitrarily large and the fitted model becomes unreliable. In contrast, Lemma 1 states that WC parameters guarantee an improvement of the true loss function no matter which is the unknown  $F_0$ , provided that  $c$  is small. This robustness is a key characteristic of WC methods.

An alternative procedure consists of minimizing  $\int Q(\theta, y)\mu(dy)$ , where  $\mu$  is a prior probability distribution for the unobservables. Again, a cautious modeller should consider a conservative prior distribution given by the solution to the problem

$$\min_{\theta \in \Theta} \max_{\mu \in \mathcal{M}} \int Q(\theta, y)\mu(dy), \quad (3)$$

where  $\mathcal{M}$  is the class of probability measures on  $\mathbb{Y}$ . The next lemma states that the estimators given by (3) are indeed WC estimators.

**Lemma 2.** Assume that  $g$  is nonnegative, then

$$\min_{\theta \in \Theta} \max_{\mu \in \mathcal{M}} \int Q(\theta, y)\mu(dy) = \min_{\theta \in \Theta} \max_{y \in \mathbb{Y}} Q(\theta, y). \quad (4)$$

The WC approach possesses a particular interest for economic decision makers, as these methods can be used to reduce the damages derived from Lucas' critique. Lucas (1976) pointed out that macro-econometric models cannot be used for policy analysis, if implementing the policy would change the conditional model in which the policy was based. Lucas argued that the fact that agents have rational expectations over future policy actions turns this situation 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  $\theta^0$ , we can use the WC parameter  $\theta^{\text{wc}}$  which is relatively robust to changes in the controls  $Y$ . A model using WC estimators could be a more stable tool for designing optimal economic policies in absence of super exogeneity.

### 3. Asymptotic properties of WC estimators

Next, we study the consistency and asymptotic normality of WC estimators. Additional notation should be introduced to derive the existence, consistency and asymptotic normality of  $\hat{\theta}_T^{\text{wc}}$ . 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,  $\mathbb{Y}(\theta^{\text{wc}})$  is the set of the WC unobservables. This set can be estimated by means of  $\mathbb{Y}_T(\hat{\theta}_T^{\text{wc}})$ , where

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

The following result summarizes some properties of min–max optimization and the WC sets.

**Lemma 3.** Let  $\Theta \subset \mathbb{R}^K$  and  $\mathbb{Y} \subset \mathbb{R}^S$  be nonempty compact sets and  $Q \in C(\Theta \times \mathbb{Y})$ . Then there exist  $\theta^{\text{wc}} \in \Theta$  and  $y^{\text{wc}} \in \mathbb{Y}$  such that

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

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 hemicontinuous and takes values that are nonempty compact convex sets. (ii) If  $Q(\theta, \cdot)$  is strictly quasi-concave on  $\mathbb{Y}$  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) quasi-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.

### 3.1. Consistency

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  $|t_T(\theta, y)| \rightarrow 0$  almost surely (in probability) when  $T \rightarrow \infty$ , uniformly in  $\theta \in \Theta$  and  $y \in \mathbb{Y}$ .

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

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

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  $\theta^{\text{wc}}$  solves the problem  $\inf_{\theta \in \Theta} \sup_{y \in \mathbb{Y}} K_{\theta^{\text{wc}}, y^{\text{wc}}}(\theta, y)$  uniquely in a neighborhood of  $\theta^{\text{wc}}$  (uniqueness is an asymptotic local identification requirement). As in the case of M estimators, local identification is a more flexible requirement than global identification. If there exist several local solutions, the modeller should choose the most convenient value according to economic literature.

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

$$\begin{aligned} Q_T(\theta, y) - Q_T(\theta^{\text{wc}}, z) &= K_{\theta^{\text{wc}}, z}(\theta, y) + t_T(\theta, y), \\ K_{\theta^{\text{wc}}, z}(\theta, y) &= Q(\theta, y) - Q(\theta^{\text{wc}}, z), \\ t_T(\theta, y) &= Q_T(\theta, y) - Q(\theta, y) + Q(\theta^{\text{wc}}, z) - Q_T(\theta^{\text{wc}}, z). \end{aligned} \quad (5)$$

Then, it is sufficient for A.1 and A.2 (and therefore the consistency of  $\widehat{\theta}_T^{\text{wc}}$ ) that  $\theta^{\text{wc}} \in \Theta$  be a locally unique solution to (1) and  $\sup_{\theta \in \Theta} \sup_{y \in \mathbb{Y}} |Q_T(\theta, y) - Q(\theta, y)| \rightarrow 0$ , 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  $\{X_i\}$ . Davidson (1994, Chapter 21), Wooldridge (1994), and Pötscher and Prucha (1997, Chapter 5) review the econometric literature, including dependent data.

**Theorem 4** (consistency of WC estimators). Let  $\widehat{\theta}_T^{\text{wc}} \in \Theta \subset \mathbb{R}^K$  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, then  $\widehat{\theta}_T^{\text{wc}} \rightarrow \theta^{\text{wc}}$  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  $\mathbb{Y}_T(\widehat{\theta}_T^{\text{wc}})$  and consider the Hausdorff distance  $d_H(A, B)$  between two nonempty Euclidean sets  $A, B$ ; i.e.,

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

where  $d(a, B) = d(B, a) = \inf_{b \in B} \|a - b\|$  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  $\mathbb{Y}_T(\widehat{\theta}_T^{\text{wc}})$  and  $\mathbb{Y}(\theta^{\text{wc}})$  are compact when  $Q_T$  and  $Q$  satisfy Statement (i) of Lemma 3.

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

A.3.  $\forall \varepsilon > 0, \exists \delta > 0,$

$$\inf_{\{y_1, y_2 \in \mathbb{Y} : \|y_1 - y_2\| > \varepsilon\}} \inf_{\theta \in \Theta} K_{\theta, y_1}(\theta, y_2) > \delta.$$

Define  $K_{\theta, y_1}(\theta, y_2) = Q(\theta, y_2) - Q(\theta, y_1)$ , as in (5). A sufficient condition for A.3 is

$$|Q(\theta, y_2) - Q(\theta, y_1)| > r(\theta) f(\|y_2 - y_1\|),$$

where  $f(x) > 0$  for all  $x > 0$ , and  $\inf_{\theta \in \Theta} r(\theta) > 0$ .

**Proposition 5** (consistency of WC unobservables). *Under assumptions A.1–A.3,*

$$d_H(\mathbb{Y}_T(\widehat{\theta}_T^{\text{wc}}), \mathbb{Y}(\theta^{\text{wc}})) \rightarrow 0,$$

almost surely (in probability).

### 3.2. Asymptotic normality

Without loss of generality, we assume that the parameter set is of the form  $\Theta = \{\theta \in \mathbb{R}^K : h(\theta) \leq 0\}$ , where  $h = (h_1, \dots, h_M)$  is a continuous vector function on  $\Theta$ . Analogously to nonlinear least-squares methods, we argue that the parametric constraints have no asymptotic effect if  $\theta^{\text{wc}}$  is an interior point of  $\Theta$ , i.e.  $\theta^{\text{wc}} \in \text{int}\{\Theta\}$  with  $\text{int}\{\Theta\} = \{\theta \in \mathbb{R}^K : h_j(\theta) < 0, j = 1, \dots, M\}$ .

To establish the asymptotic normality of WC estimators  $\widehat{\theta}_T^{\text{wc}}$ , we present the first order conditions for minimax problems, that are used to prove consistency of WC multipliers  $\{(\widehat{\mu}_i, \widehat{y}_i)\}_{i=1}^k$  associated with Problem (2). The first order necessary conditions for the solution to (1) are 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 6** (first order conditions for minimax problems). *Let  $Q : \mathbb{R}^K \times \mathbb{R}^S \rightarrow \mathbb{R}$  be  $C^1$ ,  $\Theta = \{\theta \in \mathbb{R}^K : h(\theta) \leq 0\}$ , where  $h : \mathbb{R}^K \rightarrow \mathbb{R}^P$  are  $C^1$ , and  $\mathbb{Y} \subset \mathbb{R}^S$  be a nonempty compact set. Let  $\theta^{\text{wc}}$  denote the solution to  $\min_{\theta \in \Theta} \max_{y \in \mathbb{Y}} Q(\theta, y)$ . If vectors  $\{\nabla_{\theta} h_j(\theta^{\text{wc}}) : h_j(\theta^{\text{wc}}) = 0\}$  are linearly independent, then there exist a positive integer  $k$ , vectors  $y_i \in \mathbb{Y}(\theta^{\text{wc}})$ , and multipliers  $\mu_i \geq 0$  for  $i = 1, \dots, k$ , with  $\sum_{i=1}^k \mu_i = 1$ , and  $\lambda_j \geq 0$  for  $j = 1, \dots, p$  such that*

$$\sum_{i=1}^k \mu_i \nabla_{\theta} Q(\theta^{\text{wc}}, y_i) + \sum_{j=1}^p \lambda_j \nabla_{\theta} h_j(\theta^{\text{wc}}) = 0, \quad (6)$$

$$\sum_{j=1}^p \lambda_j h_j(\theta^{\text{wc}}) = 0, \quad (7)$$

with  $1 \leq k + p \leq K + 1$ , where  $p$  is the number of nonzero  $\lambda_j$ . If  $\theta^{\text{wc}} \in \text{int}\{\Theta\}$ , Eqs. (6) and (7) simplify to  $\sum_{i=1}^k \mu_i \nabla_{\theta} Q(\theta^{\text{wc}}, y_i) = 0$ .

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

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

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

$$\begin{aligned} \int \nabla_{\theta} Q(\theta^{\text{wc}}, y)\mu(\text{d}y) + \sum_{j=1}^p \lambda_j \nabla_{\theta} h_j(\theta^{\text{wc}}) &= 0, \quad 1 - \int \mu(\text{d}y) = 0, \\ \lambda_j h_j(\theta^{\text{wc}}) &= 0, \quad h_j(\theta^{\text{wc}}) \leq 0, \quad \lambda_j \geq 0, \quad j = 1, \dots, p, \end{aligned} \quad (8)$$

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

The Lagrange multiplier  $\mu$  is a discrete measure with support in  $\mathbb{Y}(\theta^{\text{wc}})$  and can be expressed as  $\mu = \sum_{i=1}^k \mu_i \delta(y_i)$  for some  $k \leq K + 1$  and  $y_i \in \mathbb{Y}(\theta^{\text{wc}})$ . This is because the set of measures  $\mu$  satisfying the functional conditions (8) 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 (8), provided  $\{\nabla_{\theta} h_j(\theta^{\text{wc}}) : h_j(\theta^{\text{wc}}) = 0\}$  are linearly independent vectors (see Shapiro (1994) and Shapiro (1988, pp. 112–113) for details). Alternative arguments based on Caratéodory’s Theorem can be found in Hager and Presler (1987). Because  $\mu$  is a discrete probability measure, we can express (8) as

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

Any continuous minimax problem of the form (1) satisfying the assumptions in Theorem 6 can be written as

$$\min_{\rho, \theta \in \Theta} \{\rho : Q(\theta, y_i) \leq \rho, \quad i = 1, \dots, k\}. \quad (9)$$

The optima  $\theta^{\text{wc}}$  and  $\{y_i\}_{i=1}^k \subset \mathbb{Y}(\theta^{\text{wc}})$ , and Lagrange multipliers  $\{\mu_i\}$ ,  $\{\lambda_j\}$  of Problem (1) coincide with the optima and Lagrange multipliers of Problem (9). This result will be applied to prove consistency of multipliers  $\{\widehat{\mu}_i\}_{i=1}^k$  to  $\{\mu_i\}_{i=1}^k$ . To this end, the uniqueness of  $\{\mu_i, \lambda_j\}$  is also required. The following result of Shapiro (1997, Proposition 3.2) addresses this problem:

**Proposition 7** (uniqueness of  $\{\mu_i, \lambda_j\}$ ). *The following two conditions are necessary and sufficient for the uniqueness of the Lagrange multipliers measure  $\mu^{\text{wc}} = \sum_{i=1}^k \mu_i \delta(y_i)$ :*

- (1)  $\{\nabla_{\theta} Q(\theta^{\text{wc}}, y_i)\}_{i=1}^k$  are linearly independent.
- (2) For any neighborhood  $\mathcal{N}$  of the set  $\{y_i\}_{i=1}^k$  there exists a  $v$  such that  $v \cdot \nabla_{\theta} Q(\theta^{\text{wc}}, y_i) = 0$ ,  $i = 1, \dots, k$  and  $v \cdot \nabla_{\theta} Q(\theta^{\text{wc}}, y) < 0$ ,  $y \in \mathbb{Y}(\theta^{\text{wc}}) \setminus \mathcal{N}$ .

The next result gives sufficient conditions ensuring that  $\{(\widehat{\mu}_i, \widehat{y}_i)\}_{i=1}^k$  converges almost surely to the WC multipliers  $\{(\mu_i, y_i)\}_{i=1}^k$  associated with Problems (2) and (1), respectively.

**Proposition 8** (consistency of WC multipliers). *Under the assumptions in Theorem 6, Proposition 7, A.1–A.3, if, in addition,*

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

*almost surely (in probability), then  $\{(\widehat{\mu}_i, \widehat{y}_i)\}_{i=1}^k$  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.  $\theta^{\text{wc}} \in \text{int}\{\Theta\}$  solves (1), and  $\hat{\theta}_T^{\text{wc}} \rightarrow_p \theta^{\text{wc}}$ ,  
 B.2.  $\{(\hat{\mu}_i, \hat{y}_i)\}_{i=1}^{\hat{k}} \rightarrow_p \{(\mu_i, y_i)\}_{i=1}^k$ ,  
 B.3. for all  $T$ ,  $Q_T(\theta, y)$  is  $C^{2,1}$  almost surely,  $\mathbb{Y} \subset \mathbb{R}^S$  and  $\Theta \subset \mathbb{R}^K$  are nonempty compact sets, and

$$\sqrt{T} \frac{\partial Q_T(\theta^{\text{wc}}, y)}{\partial \theta} \rightarrow_d Z(y)$$

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

- B.4. for any sequence  $\hat{\theta}_T \rightarrow_p \theta^{\text{wc}}$ ,

$$\sum_{i=1}^{\hat{k}} \hat{\mu}_i \frac{\partial^2 Q_T(\hat{\theta}_T, \hat{y}_i)}{\partial \theta \partial \theta'} \rightarrow_p B := \sum_{i=1}^k \mu_i \frac{\partial^2 Q(\theta^{\text{wc}}, y_i)}{\partial \theta \partial \theta'},$$

where  $B$  is a nonsingular deterministic real matrix.

**Theorem 9** (asymptotic normality). *Let  $\hat{\theta}_T^{\text{wc}}$  be the solution to (2). Assume B.1–B.4. Then,  $\sqrt{T}(\hat{\theta}_T^{\text{wc}} - \theta^{\text{wc}}) \rightarrow_d N(0, B^{-1}AB^{-1})$ , where  $A = \sum_{i=1}^k \sum_{j=1}^k \mu_i \mu_j R(y_i, y_j)$  is a positive definite real matrix.*

Relaxing the compactness condition of  $\mathbb{Y}$  in Theorem 9, asymptotic normality can be analogously proved, considering Eq. (8) instead of Eqs. (6) and (7). But the estimation of the measure  $\mu$  becomes more involved (see the concluding remarks).

Consistency of WC estimators and multipliers (considered in Assumptions B.1 and B.2) can be proven using Theorem 4 and Propositions 5 and 8. Assumption B.3 can be established by 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

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

uniformly on  $C(\mathbb{Y} \times \Theta)$ , which requires a ULLN. In the appendix we provide an alternative sufficient condition for B.4 (see Proposition 12).

Often,

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

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

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

and  $B$  by

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



For dependence cases, we can consider analogous expressions based on the spectral density evaluated at zero, such as the Barlett–Newey–West estimator.

### 3.3. Constrained estimation and testing

Theorem 9 is useful when  $\theta^{\text{wc}} \in \text{int}\{\Theta\}$ . But often, the parametric set is defined to include some equality constraints, meaning that  $\theta^{\text{wc}}$  is not an interior point. In this section we consider the case where  $h_j(\theta^{\text{wc}}) = 0$  for  $j = 1, \dots, p$ , and these constrains are used in the estimation process. We define

$$\Theta^{\text{null}} = \{h_j(\theta) = 0, j = 1, \dots, p\} \cap \{h_j(\theta) < 0, j = p + 1, \dots, M\},$$

and the constrained WC estimator (CWC) of  $\theta^{\text{wc}}$  given by

$$\widehat{\theta}_T^{\text{cwc}} = \arg \min_{\theta \in \Theta^{\text{null}}} \max_{y \in \mathbb{Y}} Q_T(\theta, y).$$

Let us denote the Lagrange multipliers associated to the constrained minimax problem by  $\widehat{\lambda}_T^{\text{cwc}}$ . We obtain the asymptotic distribution of  $\sqrt{T}(\widehat{\theta}_T^{\text{cwc}} - \theta^{\text{wc}}, \widehat{\lambda}_T^{\text{cwc}})$ , which allows us to derive asymptotic parametric tests. The proof of asymptotic normality is similar to that of Theorem 9; however, we should slightly modify Assumption B.1 as follows:

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

**Theorem 10** (asymptotic normality of constrained WC estimators). Let  $\widehat{\theta}_T^{\text{cwc}}$  be the solution to (2) with Lagrange multipliers  $\widehat{\lambda}_T^{\text{cwc}}$ . Assume B.1', B.2, B.3 and B.4. Then,

$$\sqrt{T} \begin{pmatrix} \widehat{\theta}_T^{\text{cwc}} - \theta^{\text{wc}} \\ \widehat{\lambda}_T^{\text{cwc}} \end{pmatrix} \rightarrow_d N(0, V),$$

where

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

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

$$C_{11} = B^{-1} - B^{-1}H'(HB^{-1}H')^{-1}HB^{-1},$$

$$C_{12} = (HB^{-1}H')^{-1}HB^{-1}.$$

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

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

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  $B^{-1}H'(HB^{-1}H')^{-1}HB^{-1}$  is nonnegative definite). It means that by incorporating valid restrictions we cannot reduce efficiency, but generally improve it.

Theorem 10 can be used to test Lagrange multiplier hypotheses. For example, the statistic for testing  $H_0 : h_j(\theta^{\text{wc}}) = 0$  for  $j = 1, \dots, p$ , is  $\Upsilon_T := T \widehat{\lambda}' \widehat{V}_{22}^{-1} \widehat{\lambda} \rightarrow_d \chi_{K-p}^2$ , where  $\widehat{V}_{22} \rightarrow_p V_{22} > 0$ . Other asymptotic tests, such as Wald-type tests and generalized likelihood ratio tests, can be derived in a similar way, using  $\sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \nabla_{\theta} Q_T(\theta, \widehat{y}_i)$  as a score function.

#### 4. WC estimation for overidentified models

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

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

where  $W_T \rightarrow_p W$ . A review of the literature can be found e.g., in Wooldridge (1994).

This section shows how the WC approach is embedded in the GMM framework. Assume that  $Y$  is unobserved and consider the loss function  $Q(\theta, y) = E[g(X, y, \theta)]' W E[g(X, y, \theta)]$  on  $\Theta \times \mathbb{Y}$ . We define  $\theta^{\text{wc}}$  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

$$Q_T(\theta, y) = \left( T^{-1} \sum_{t=1}^T g(X_t, y, \theta) \right)' W_T \left( T^{-1} \sum_{t=1}^T g(X_t, y, \theta) \right), \quad (10)$$

where  $W_T \rightarrow_p W$  almost surely (in probability), the WC-GMM estimator  $\widehat{\theta}_T^{\text{wc}}$  of  $\theta^{\text{wc}}$  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 4 and Proposition 5). However, as it is usually done for the classical GMM method, the asymptotic normality for WC-GMM can be established using weaker conditions than those considered in Theorem 9. The following assumptions are introduced:

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

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

$W$  is positive definite, and  $\widehat{\theta}_T^{\text{wc}} \rightarrow_p \theta^{\text{wc}}$ .

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

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

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

D.4.

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

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

**Theorem 11** (asymptotic normality of WC GMM). Let  $\widehat{\theta}_T^{\text{wc}}$  be the solution to (2), and  $Q_T$  be given by (10). Assume D.1, B.2, D.3, and D.4. Then,

$$\sqrt{T}(\widehat{\theta}_T^{\text{wc}} - \theta^{\text{wc}}) \rightarrow_d N(0, B^{-1}AB^{-1}),$$

where

$$A = \sum_{i=1}^k \sum_{j=1}^k \mu_i \mu_j S(y_i, \theta^{\text{wc}}) W R(y_i, y_j) W' S(y_j, \theta^{\text{wc}})',$$

$$B = \sum_{i=1}^k \mu_i S(y_i, \theta^{\text{wc}}) W S(y_i, \theta^{\text{wc}})'.$$

The asymptotic variance of WC–GMM estimators is more complex than the one of the classical GMM estimators. If  $R(y_i, y_j) = S(y_i, \theta^{\text{wc}}) S(y_j, \theta^{\text{wc}})'$ , the asymptotic variance is  $B^{-1} A B^{-1} = I$ . In the WC context, if  $k > 1$  it is not straightforward to ensure that  $A = B^{-1}$  by an appropriate choice of  $W$ . The asymptotic distribution for WC–GMM constrained estimators can be derived analogously to Theorem 10.

GMM is the benchmark approach for overidentified models, but other methods have been considered in the literature and WC ideas can be adapted to their premises. For instance, some authors consider infinite moment conditions  $E[g_s(X, Y, \theta)] = 0$  indexed by elements  $s$  in a Euclidean space, and the parameter  $\theta^0$  is identified as a minimizer of  $Q(\theta) = \int |E[g_s(X, Y, \theta)]|^2 \mu(ds)$  for some weight measure  $\mu$  (see e.g., [Domiguez and Lobato, 2004](#)). If the variable  $Y$  is unobserved, the WC approach can be extended to this context considering

$$Q(\theta, y) = \int |E[g_s(X, y, \theta)]|^2 \mu(ds).$$

## 5. Computational issues

To obtain WC estimators, we are faced with the problem of solving a minimax continuous problem. We consider the global optimization algorithm developed by [Shimizu and Aiyoshi \(1980\)](#) and [Žaković and Rustem \(2003\)](#), see also the monograph of [Rustem and Howe \(2002\)](#). They consider an algorithm for solving semi-infinite programming problems, as any continuous minimax problem of the form  $\min_{\theta \in \Theta} \max_{y \in \mathbb{Y}} Q_T(\theta, y)$  can be written as  $\min_{\theta \in \Theta, \rho \in \mathbb{R}} \{\rho : \max_{y \in \mathbb{Y}} Q_T(\theta, y) \leq \rho\}$ , which is equivalent to the semi-infinite problem:

$$\begin{aligned} \min_{\theta \in \Theta, \rho \in \mathbb{R}} \quad & \rho \\ \text{s.t.} \quad & Q_T(\theta, y) \leq \rho \quad \text{for all } y \in \mathbb{Y}. \end{aligned} \tag{11}$$

For an introduction to semi-infinite programming, see [Hettich and Kortanek \(1993\)](#) and [Reemtsen and Görner \(1998\)](#).

The [Žaković and Rustem \(2003\)](#) 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:

$$\min_{\theta^{l+1} \in \Theta, \rho^{l+1} \in \mathbb{R}} \{\rho^{l+1} : Q_T(\theta^{l+1}, y_i) \leq \rho^{l+1}, i = 1, \dots, k_l\}, \tag{12}$$

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

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

iterate, otherwise if  $\max_{y \in \mathbb{Y}} Q_T(\theta^{l+1}, y) \leq \rho^{l+1} + \varepsilon$ , terminate and  $\widehat{\theta}_T^{\text{wc}} = \theta^{l+1}$  is a solution of the minimax problem. This algorithm terminates in a finite number of iterations. Under convexity assumptions of Problem (11), the Lagrange multipliers  $\{\mu_i\}_{i=1}^k$  associated with the last iteration of Problem (12) are the coefficients  $\{\mu_i\}_{i=1}^k$  in Theorem 6.

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  $\{y_i\}_{i=1}^{k_l} \subset \mathbb{Y}_T(\theta^l)$  for all  $\theta^l \in \Theta$  by computing the global maximizers in the program  $\max_{y \in \mathbb{Y}} Q_T(\theta^l, y)$ . In global optimization algorithms, all candidates for local maximizers must usually be bracketed by a comparison of function values  $Q_T(\theta^l, y)$  on a sufficiently dense finite subset of  $\mathbb{Y}$ . To reduce the cost of computing global optima, it is recommended that the domains  $\Theta$  and  $\mathbb{Y}$  be restricted as much as possible given the information available.

Table 1  
Finite sample results for WC in model (i) with  $\kappa = 0$

	LS	Omitting	Prior $N$	Prior $U$	WC
$E[\hat{\alpha}]$	-2.05	0.2	1.67	-1.9	-2.28
	-2.06	-1.98	-2.11	-2.05	-2.19
$V[\hat{\alpha}]$	3.58	20.3	19.3	6.3	4.65
	3.46	20.2	15.31	4.78	4.27
$MSE(\hat{\alpha})$	3.59	25.1	19.48	6.3	4.73
	3.47	20.2	15.32	4.78	4.3
$E[\hat{\beta}]$	2.01	—	0.07	3.09	1.01
	1.9	—	0.003	3.11	1.01
$V[\hat{\beta}]$	0.02	—	0.008	0.09	0.05
	0.03	—	0.0079	0.09	0.05
$MSE(\hat{\beta})$	0.02	—	3.98	1.29	0.96
	0.03	—	3.99	1.34	0.97
$Q(\hat{\alpha}, \hat{\beta})$	0.92	5.6	5.5	2.58	2.03
	0.93	5.5	5.6	2.57	2.12
$V(\hat{\alpha}, \hat{\beta})$	0.05	1.0	1.1	0.71	0.2
	0.05	1.0	0.94	0.52	0.16

We have implemented the algorithm using MATLAB 6.5 on an Intel Centrino Pentium M 1.6 GHz with machine precision  $10^{-16}$ . Each problem (11) of the Monte Carlo studies in Section 5 and the financial application in Section 6 have been computed using the MATLAB subroutine `fseminf` corresponding to the *Optimization toolbox*. This routine is suited for optimization problems with a semi-infinitely constrained multivariable nonlinear function.

## 6. Monte Carlo study of finite sample behavior

A Monte Carlo study was conducted in order to study the finite sample performance of WC estimates. We first consider the linear regression model (i)  $y_t = \alpha x_{1t} + \beta x_{2t} + u_t$ , where  $(\alpha, \beta)$  denote the two-dimensional parameter vector,  $x_{1t} \sim N(0, 1)$  and  $x_{2t} = \kappa x_{1t} + \varepsilon_t$ , where  $\varepsilon_t \sim \exp\{N(0, 0.3)\}$  and  $u_t \sim N(0, 1)$  are identically and independently distributed random variables, for all  $t = 1, \dots, T$ ; the expression  $N(0, 1)$  denotes the standard normal distribution. Assuming  $\alpha^0 = -2$  and  $\beta^0 = 2$ , the experiment was carried out for  $T = 30$  and  $40$ . We also consider different levels of dependence between  $x_{1t}$  and  $x_{2t}$ , setting  $\kappa = 0$  (independence),  $\kappa = 0.3$  and  $0.5$ .

Consider the problem of estimating the WC estimators for  $(\alpha, \beta)$  assuming that  $x_2$  is unobservable. Following the least-square approach, we define  $Q_T(\alpha, \beta, x_2) = T^{-1} \sum_{t=1}^T g(y_t, x_{1t}, x_2, \alpha, \beta)$ , with  $g(y, x_1, x_2, \alpha, \beta) = (y - (\alpha x_1 + \beta x_2))^2$  and the WC problem as

$$\min_{\alpha \in [-5.0], \beta \in [0.3]} \max_{x_2 \in [0.3]} Q_T(\alpha, \beta, x_2).$$

In order to illustrate the accuracy of the asymptotic distribution, we perform a Monte Carlo with  $N=400$  realizations. We have also considered other estimators, such as: (1) the least-squares approach assuming that both  $x_1, x_2$  are observable; (2) the least-squares estimator of  $\alpha$  omitting  $x_2$ ; (3) the estimator which minimizes

$$T^{-1} \sum_{t=1}^T \int g(y_t, x_{1t}, x_2, \alpha, \beta) f(x_2) dx_2$$

in  $\alpha, \beta$  where  $f$  is a  $N(0, 1)$  prior density; and (4) analogous to the previous estimator with  $f$  a  $U(0, 1)$  prior density. All these estimators are compared with the WC estimator.

For independent regressors,  $\kappa = 0$ , Table 1 reports expectation, variance and mean square error of the considered estimators with respect to true parameters, obtained from the Monte Carlo simulation. Each cell contains two values, the upper value refers to estimators with  $T = 30$  and below it is the value computed for  $T = 40$ . Table 1 also reports the numerical value of the loss function  $Q(\hat{\alpha}, \hat{\beta}) = E[T^{-1} \sum_{t=1}^T g(y_t, x_{1t}, x_{2t}, \hat{\alpha}, \hat{\beta})]$  and its variance

Table 2  
Finite sample results for WC in model (i) with  $\kappa = 0.3$  and  $0.5$

	$\kappa = 0.3$					$\kappa = 0.5$				
	LS	Omitting	Prior $N$	Prior $U$	WC	LS	Omitting	Prior $N$	Prior $U$	WC
$E[\widehat{\alpha}]$	-1.97	-0.42	-1.3	-1.5	-2.12	-2.07	-1.0	-0.71	-0.99	-1.84
	-2.07	-1.47	-1.44	-1.44	-1.84	-2.22	-0.87	-0.99	-1.06	-1.89
$V[\widehat{\alpha}]$	3.86	23.15	24.03	6.26	4.23	3.6	20.49	22.2	5.83	4.24
	3.31	18.8	17.53	4.37	4.1	3.99	20.03	16.41	4.8	4.09
$MSE(\widehat{\alpha})$	3.86	25.65	24.5	6.51	4.24	3.6	21.49	23.86	6.85	4.026
	3.32	19.08	17.83	4.68	4.12	4.04	21.28	17.42	5.69	4.1
$E[\widehat{\beta}]$	2.01	—	0.004	3.1	1.01	2.0	—	0.005	3.11	1.01
	2.0	—	-0.008	3.12	1.02	2.0	—	0.003	3.15	1.02
$V[\widehat{\beta}]$	0.02	—	0.008	0.11	0.008	0.02	—	0.007	0.10	0.006
	0.03	—	0.008	0.08	0.005	0.02	—	0.008	0.08	0.007
$MSE(\widehat{\beta})$	0.02	—	3.99	1.34	0.96	0.02	—	3.98	1.35	0.96
	0.03	—	4.04	1.35	0.96	0.02	—	3.99	1.4	0.95
$Q(\widehat{\alpha}, \widehat{\beta})$	0.96	5.7	5.47	2.64	2.09	0.94	5.57	5.59	2.64	2.08
	0.9	5.6	5.66	2.6	2.14	0.94	5.69	5.63	2.65	2.06
$V(\widehat{\alpha}, \widehat{\beta})$	0.06	1.08	1.24	0.78	0.23	0.06	0.95	1.13	0.74	0.22
	0.05	0.98	0.92	0.62	0.18	0.06	1.12	1.07	0.58	0.15

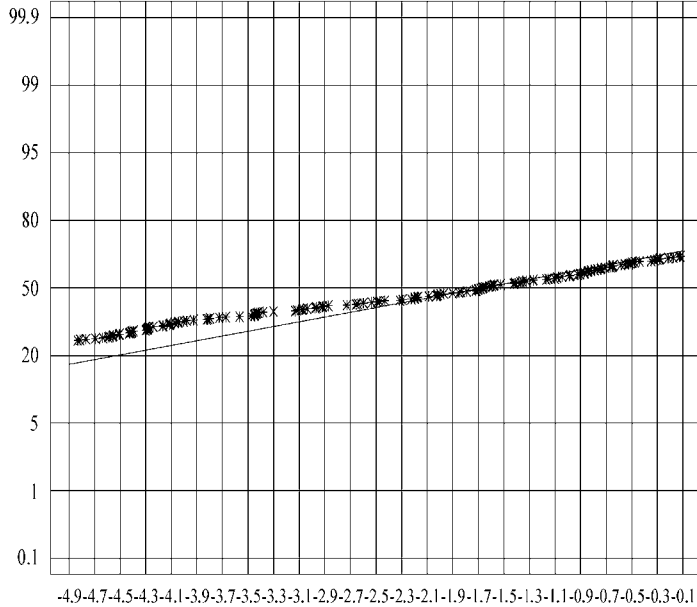


Fig. 1. Normal probability plot for  $\widehat{\alpha}^{wc}$  of model (i), with  $T = 30$  and  $\kappa = 0$ .

$V(\widehat{\alpha}, \widehat{\beta}) = \text{Var}[T^{-1} \sum_{t=1}^T g(y_t, x_{1t}, x_{2t}, \widehat{\alpha}, \widehat{\beta})]$ , computed from the Monte Carlo simulation for each of these estimators. From Table 1, it can be deduced that the WC approach is a robust strategy in the presence of unobservables, as documented in Lemma 1. The case of dependent regressors is considered in Table 2, showing an analogous pattern.

Figs. 1 and 2 display the normal probability plot for  $\widehat{\alpha}^{wc}$  and  $\widehat{\beta}^{wc}$ , respectively, for  $T = 30$  and  $\kappa = 0$ .

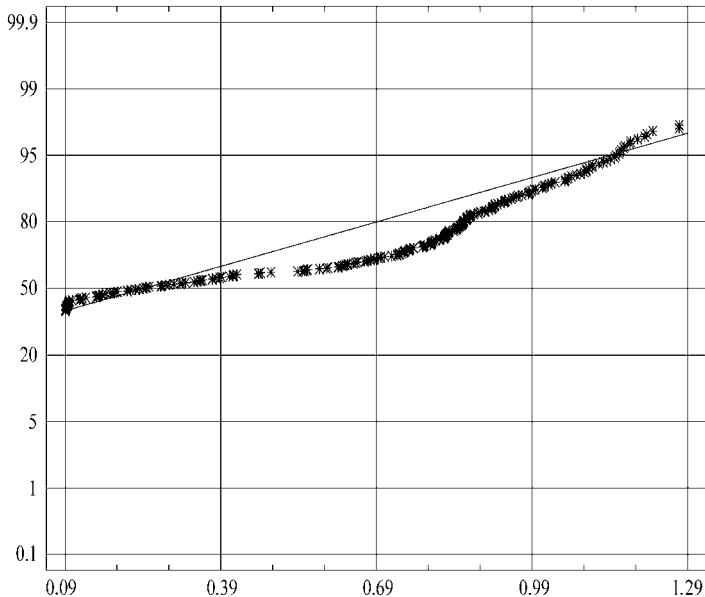


Fig. 2. Normal probability plot for  $\widehat{\beta}^{\text{wc}}$  of model (i), with  $T = 30$  and  $\kappa = 0$ .

Figs. 1 and 2 show that the normal approximation is satisfactory for  $T = 30$ , although there is room for improvement. 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 for small samples. But for large samples they may not be worth emphasizing over first order weak asymptotic approximations, as in classical M-estimation.

A nonlinear regression model (ii)  $y_t = \exp(\alpha x_{1t} + \beta x_{2t}) + u_t$  was also simulated, where  $(\alpha, \beta)$  denote the two-dimensional parameter vector,  $u_t \sim N(0, 1)$ , and the regressors  $x_{1t}$  and  $x_{2t}$  are defined as in the model (i); i.e.,  $x_{1t} \sim N(0, 1)$  and  $x_{2t} = \kappa x_{1t} + \varepsilon_t$  with  $\varepsilon_t \sim \exp\{N(0, 0.3)\}$ . Assuming  $\alpha^0 = 0.75$ , and  $\beta^0 = 0.2$ , a Monte Carlo with  $N = 400$  realizations was carried out for  $T = 30$  and  $40$ , and for the values  $\kappa = 0, 0.3$  and  $0.5$ . We consider the problem of estimating the WC estimators for  $(\alpha, \beta)$  assuming that  $x_2$  is unobservable. We define  $Q_T(\alpha, \beta, x_2) = T^{-1} \sum_{t=1}^T g(y_t, x_{1t}, x_2, \alpha, \beta)$ , with  $g(y, x_1, x_2, \alpha, \beta) = (y - \exp(\alpha x_1 + \beta x_2))^2$ , and the WC problem as

$$\min_{\alpha \in [-3, 4], \beta \in [-1, 1]} \max_{x_2 \in [0, 2]} Q_T(\alpha, \beta, x_2).$$

Analogously to the previous example, we have also considered the estimators: (1) the nonlinear least-squares approach assuming that both  $x_1, x_2$  are observable; (2) the nonlinear least-squares estimator of  $\alpha$  omitting  $x_2$ ; (3) the estimator based on a prior  $N(0, 1)$  density, as in the previous experiment; and (4) analogous to the previous case with a prior  $U(0, 2)$  density. These estimators are compared with WC estimator.

Table 3 reports expectation, variance and mean square error of the considered estimators with respect to true parameters, obtained from the Monte Carlo simulation for independent regressors,  $\kappa = 0$ . Each cell contains two values, the upper value refers to estimators with  $T = 30$  and below is the value computed for  $T = 40$ . Table 3 also reports the numerical value of the loss function  $Q(\widehat{\alpha}, \widehat{\beta}) = E[T^{-1} \sum_{t=1}^T g(y_t, x_{1t}, x_{2t}, \widehat{\alpha}, \widehat{\beta})]$  and its variance  $V(\widehat{\alpha}, \widehat{\beta}) = \text{Var}[T^{-1} \sum_{t=1}^T g(y_t, x_{1t}, x_{2t}, \widehat{\alpha}, \widehat{\beta})]$  computed from the Monte Carlo simulation for each of these estimators. The case of dependent regressors,  $\kappa = 0.3$  and  $0.5$ , is considered in Table 4. From Tables 3 and 4 it can be deduced that WC estimators guarantee an improvement of the true loss function and the loss associated to the WC estimators and its variability are usually lower than the ones obtained by omitting the unobservable or considering arbitrary priors.

Table 3  
Finite sample results for model (ii) with  $\kappa = 0$ .

	LS	Omitting	Prior $N$	Prior $U$	WC
$E[\widehat{\alpha}]$	0.71	1.02	1.08	0.67	0.97
	0.67	0.94	1.05	0.79	1.1
$V[\widehat{\alpha}]$	1.83	3.48	4.78	2.77	3.19
	1.86	3.42	3.9	2.09	2.4
$MSE(\widehat{\alpha})$	1.83	3.55	4.89	2.77	3.24
	1.87	3.45	4.03	2.1	2.55
$E[\widehat{\beta}]$	0.18	—	0.001	0.13	0.14
	0.19	—	0.0003	0.13	0.15
$V[\widehat{\beta}]$	0.016	—	0.0002	0.013	0.007
	0.013	—	0.0002	0.009	0.006
$MSE(\widehat{\beta})$	0.016	—	0.03	0.017	0.01
	0.014	—	0.04	0.013	0.007
$Q(\widehat{\alpha}, \widehat{\beta})$	2.05	2.52	2.45	2.17	2.2
	2.04	2.51	2.55	2.17	2.19
$V(\widehat{\alpha}, \widehat{\beta})$	0.06	0.2	0.25	0.08	0.15
	0.05	0.19	0.20	0.007	0.11

Table 4  
Finite sample results for WC in model (ii) with  $\kappa = 0.3$  and 0.5

	$\kappa = 0.3$					$\kappa = 0.5$				
	LS	Omitting	Prior $N$	Prior $U$	WC	LS	Omitting	Prior $N$	Prior $U$	WC
$E[\widehat{\alpha}]$	0.76	1.01	1.01	1.01	1.01	0.73	1.05	1.04	1.05	1.04
	0.76	1.01	1.01	1.008	1.01	0.74	1.04	1.05	1.04	1.04
$V[\widehat{\alpha}]$	0.06	0.003	0.003	0.0048	0.003	0.10	0.0035	0.036	0.004	0.004
	0.05	0.004	0.003	0.003	0.003	0.05	0.004	0.002	0.003	0.003
$MSE(\widehat{\alpha})$	0.06	0.07	0.07	0.07	0.07	0.10	0.09	0.09	0.09	0.09
	0.06	0.07	0.07	0.07	0.07	0.05	0.09	0.09	0.09	0.09
$E[\widehat{\beta}]$	0.19	—	$1.4 \times 10^{-5}$	0.004	0.002	0.2	—	$-3.2 \times 10^{-6}$	0.006	0.009
	0.19	—	$-7.5 \times 10^{-6}$	0.006	0.0002	0.2	—	$9.1 \times 10^{-6}$	0.006	0.007
$V[\widehat{\beta}]$	0.03	—	$1.1 \times 10^{-7}$	0.003	0.001	0.04	—	$1.01 \times 10^{-7}$	$3 \times 10^{-4}$	0.002
	0.03	—	$6.3 \times 10^{-8}$	$2 \times 10^{-4}$	$8 \times 10^{-4}$	0.02	—	$8 \times 10^{-8}$	0.002	0.001
$MSE(\widehat{\beta})$	0.03	—	0.039	0.038	0.04	0.04	—	0.04	0.03	0.038
	0.03	—	0.04	0.03	0.04	0.02	—	0.039	0.03	0.03
$Q(\widehat{\alpha}, \widehat{\beta})$	4.18	4.226	4.19	4.2	4.2	4.44	4.49	4.45	4.52	4.47
	4.2	4.24	4.23	4.2	4.23	4.44	4.487	4.49	4.45	4.47
$V(\widehat{\alpha}, \widehat{\beta})$	0.24	0.257	0.27	0.29	0.254	0.26	0.27	0.28	0.28	0.26
	0.21	0.29	0.23	0.2	0.21	0.22	0.31	0.23	0.19	0.21

For model (ii) we have considered the effect of the interval for the unobservable on the WC estimators, using Monte Carlo simulation. Table 5 shows that  $\widehat{\alpha}^{\text{wc}}$  and  $\widehat{\beta}^{\text{wc}}$  do not suffer too strongly from an arbitrary selection of the interval for the nonnegative variable  $x_2$ .

## 7. A financial application in a GMM context

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

Table 5

Sensitivity analysis of WC with respect to  $\mathbb{Y}$  in model (ii), with  $\kappa = 0.3$  and  $T = 40$ 

$Y$	[0, 1]	[0, 2]	[0, 3]	[0, 4]
$E[\widehat{\alpha}^{\text{wc}}]$	1.02	1.01	1.01	1.01
$V[\widehat{\alpha}^{\text{wc}}]$	0.003	0.003	0.003	0.002
$MSE(\widehat{\alpha}^{\text{wc}})$	0.07	0.07	0.07	0.07
$E[\widehat{\beta}^{\text{wc}}]$	-0.007	0.0002	-0.002	0.0009
$V[\widehat{\beta}^{\text{wc}}]$	0.001	$8 \times 10^{-4}$	0.002	$2.2 \times 10^{-5}$
$MSE(\widehat{\beta}^{\text{wc}})$	0.04	0.04	0.04	0.039
$Q(\widehat{\alpha}^{\text{wc}}, \widehat{\beta}^{\text{wc}})$	4.27	4.23	4.21	4.21
$V(\widehat{\alpha}^{\text{wc}}, \widehat{\beta}^{\text{wc}})$	0.2	0.21	0.25	0.23

investment, solving the dynamic optimization problem:

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

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,

$$p_{j,t} u'(c_t) = \theta_1^{m_j} E[r_{j,t+m_j} u'(c_{t+m_j}) | \mathcal{F}_t] \Leftrightarrow$$

$$0 = E \left[ \left( \theta_1^{m_j} \frac{u'(c_{t+m_j})}{u'(c_t)} \frac{r_{j,t+m_j}}{p_{j,t}} - 1 \right) \middle| \mathcal{F}_t \right], \quad j = 1, \dots, N.$$

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

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

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

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

Following Hansen and Singleton (1982), this expression can be used to estimate by GMM when all the required information is available.

In this financial example, Hansen and Singleton (1982) 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 WC approach 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 WC GMM estimation associated with the moment conditions

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



with  $X_t = (c_t, c_{t+m_j}, r_{j,t+m_j}, Z_t)$ . 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  $t = 2, \dots, T-1$ .

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 WC parameters estimates obtained using the described procedure are  $\hat{\theta}_1^{\text{wc}} = 0.85$  and  $\hat{\theta}_2^{\text{wc}} = 1.25$ , the maximum optimum in  $\mathbb{Y}_T(\hat{\theta})$  is  $\hat{y}_1 = 155$ , the associated Lagrange multiplier is  $\hat{\mu}_1 = -1$ , and the standard deviations are  $sd(\hat{\theta}_1^{\text{wc}}) = 0.2$  and  $sd(\hat{\theta}_2^{\text{wc}}) = 0.12$ , respectively.

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.

## 8. Concluding remarks

This paper proposes a WC or minimax approach for estimating econometric models containing unobservable components. Instead of postulating parameters or distributional assumptions for the unobservable variables, we consider the estimators that best fit the available data in view of the worst realization of the unobservable. This method is robust with respect to the unknown probability distribution of unobservables. Computing WC estimators involves solving a minimax continuous problem, which is usually analytically intractable, and the use of efficient numerical methods is required. In particular, we use the global optimization algorithm considered by Žaković and Rustem (2003). The numerical results of Monte Carlo simulations and the financial illustration reveals that the proposed estimation method should be seen as an effective complement to other available methodologies with unobservables.

WC estimators can be considered for other problems rather than unobservable data. For example:

- when the specification of alternative nested models is considered. Let  $g_1(X_t, \theta), \dots, g_S(X_t, \theta)$  denote alternative log likelihood models. Then estimators can be constructed solving

$$\min_{\theta \in \Theta} \max_{y \in \mathbb{Y}} \left\{ \frac{1}{T} \sum_{t=1}^T y_j g_j(X_t, \theta) \right\},$$

where  $\mathbb{Y} = \{y \in \mathbb{R}^S : \sum_{j=1}^S y_j = 0, y_j \geq 0\}$  is the positive simplex in  $\mathbb{R}^S$ ;

- in overidentified models with moment conditions  $E[g_j(X_t, \theta^0)] = 0$  for  $j = 1, \dots, S$  and  $\Theta \subset \mathbb{R}^K$  with  $S > K$ . When  $S$  is very large, the optimal GMM may have a high bias in small samples. To improve it, the literature has considered a particular type of WC estimators known as generalized empirical likelihood (GEL) estimators, given by

$$\arg \min_{\theta \in \Theta} \max_{y \in \mathbb{Y}} \left\{ \frac{1}{T} \sum_{t=1}^T \rho \left( \sum_{j=1}^S y_j g_j(X_t, \theta) \right) \right\},$$

where  $\mathbb{Y}$  is the positive simplex,  $\rho$  an appropriate function and  $\theta^0 = \theta^{\text{wc}}$ . GEL estimators have smaller high order bias and often smaller high order variance than GMM estimators; see Newey and Smith (2004) and Newey et al. (2005).

These two applications show that the boundedness of  $\mathbb{Y}$  is often satisfied in practice. However, there may exist situations in which it would be useful to relax this hypothesis and Lemma 2 can be used to overcome this issue.

Let  $\mu^{\text{wc}}$  be a unique solution of problem (4) in Lemma 2.

- (1) If  $\mu^{\text{wc}}$  is absolutely continuous with density function  $m^{\text{wc}}(y)$ , consider a Schauder basis  $\{\phi_k\}$  for the  $L_1$  space (although we may also assume that the density function belongs to  $L_2$  and set an orthonormal basis), then  $m^{\text{wc}}(y) = \sum_k c_k \phi_k(y)$  and we can estimate the solution to  $\min_{\theta} \max_{\mu} \int Q(\theta, y) \mu(dy)$  by the solution to

$$\min_{\theta \in \Theta} \max_{\{c_1, \dots, c_K\}} \sum_{k=1}^K c_k \int Q_T(\theta, y) \phi_k(y) dy.$$

Often, a large but finite number of basis functions is enough to get a good approximation, and we can use the asymptotic results of this paper. Alternatively, if  $K \rightarrow \infty$  when  $T \rightarrow \infty$  it can be seen as a sieves estimation problem, but this case lies beyond the scope of this paper.

- (2) If  $\mu^{\text{wc}}$  has a finite number  $K$  of jumps, and we can estimate the solution to  $\min_{\theta} \max_{\mu} \int Q(\theta, y) \mu(dy)$  by the solution to

$$\min_{\theta \in \Theta} \max_{\{\mu_i, y_i\}} \sum_{i=1}^K \mu_i \int Q_T(\theta, y_i).$$

Nonetheless, the boundedness of  $\mathbb{Y}$  is in fact a condition that can be fully assumed in practice. Notice that economic variables are usually bounded; e.g. variables related to wealth, human or natural resources which are scarce and limited. Further, a practitioner can always establish the range of values of the unobservable with probability almost one from empirical work. This range could come from market agents or economic literature.

## Acknowledgments

We thank Prof. B. Rustem, M. Dominguez and C. Velasco and four anonymous referees of this journal for their helpful comments and suggestions, which have led to an improved version of this paper. This research has been partly supported by the European Commission, through projects FP6-2004-505509, FP6-2004-505469, the Ministerio de Educación y Ciencia of Spain, through the project SEJ2004-00672, and Comunidad Autónoma de Madrid, through the project S-0505/TIC/000230.

## Appendix: Proofs

**Proof of Lemma 1.** Let  $\mathbb{X}$  denote the support of  $X$ . Under H.0, for all  $\theta \in \Theta$ ,

$$\begin{aligned} Q(\theta) &= E[g(X, Y, \theta)] = \int_{\mathbb{X} \times \mathbb{Y}} g(x, y, \theta) F_0(dx, dy) = \int_{\mathbb{Y}} \left( \int_{\mathbb{X}} g(x, y, \theta) f_0(y|x) F_0(dx) \right) dy \\ &\leq \mu(\mathbb{Y}) \times \max_{y \in \mathbb{Y}} \int_{\mathbb{X}} g(x, y, \theta) f_0(y|x) F_0(dx) \leq c \max_{y \in \mathbb{Y}} \int_{\mathbb{X}} g(x, y, \theta) F_0(dx) = c \max_{y \in \mathbb{Y}} Q(\theta, y). \end{aligned}$$

Under H.0',

$$\begin{aligned} Q(\theta) &= \int_{\mathbb{X} \times \mathbb{Y}} g(x, y, \theta) F_0(dx, dy) \leq c \int_{\mathbb{Y}} \left\{ \int_{\mathbb{X}} g(x, y, \theta) F_0(dx) \right\} F_0(dy) \\ &\leq c \max_{y \in \mathbb{Y}} \int_{\mathbb{X}} g(x, y, \theta) F_0(dx) = c \max_{y \in \mathbb{Y}} Q(\theta, y). \quad \square \end{aligned}$$

**Proof of Lemma 2.** For all  $\theta \in \Theta$  and  $\mu \in \mathcal{M}$ ,

$$\int Q(\theta, y) \mu(dy) \leq \max_{y \in \mathbb{Y}} Q(\theta, y) \int \mu(dy) = \max_{y \in \mathbb{Y}} Q(\theta, y).$$

Furthermore, for all  $\theta \in \Theta$ , it is satisfied that

$$\max_{\mu \in \mathcal{M}} \int Q(\theta, y) \mu(dy) \geq \int Q(\theta, y) \mu^{\text{wc}}(dy) = \max_{y \in \mathbb{Y}} Q(\theta, y),$$

where  $\mu^{\text{wc}}$  is any measure of probability with all its mass in

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

**Proof of Theorem 4.** 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  $B_\varepsilon = \{\theta \in \Theta, \|\theta - \theta^{\text{wc}}\| < \varepsilon\}$  and  $B_\varepsilon^c = \Theta \setminus B_\varepsilon$ . Then, as  $\theta^{\text{wc}} \in B_\varepsilon$

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

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

**Proof of Proposition 5.** We will prove that  $\sup_{y^{\text{wc}} \in \mathbb{Y}(\theta^{\text{wc}})} d(\mathbb{Y}_T(\widehat{\theta}_T^{\text{wc}}), y^{\text{wc}}) \rightarrow 0$ , for any  $y^{\text{wc}} \in \mathbb{Y}(\theta^{\text{wc}})$  in Part 1. Part 2 proves the proposition.

*Part 1.* For any  $y^{\text{wc}} \in \mathbb{Y}(\theta^{\text{wc}})$  and any  $\varepsilon > 0$ , we define the set  $N_\varepsilon(y^{\text{wc}}) = \{y \in \mathbb{Y} : \|y - y^{\text{wc}}\| < \varepsilon\}$ , and  $N_\varepsilon(y^{\text{wc}})^c$  its complementary. Since  $y^{\text{wc}} \in N_\varepsilon(y^{\text{wc}})$ , then

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

Therefore, under A.1–A.3,

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

and the result follows.

*Part 2.* Note that  $\{d_H(\mathbb{Y}_T(\widehat{\theta}_T^{\text{wc}}), \mathbb{Y}(\theta^{\text{wc}})) > \varepsilon\}$  is equal to

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

We have proved that the first set union in the right-hand side is included in  $\{\sup_{\theta \in \Theta} \sup_{y \in \mathbb{Y}} |t_T(\theta, y)| > \delta\}$ . Next we consider the second union of sets. Let  $N_\varepsilon(\widehat{y}) = \{y \in \mathbb{Y} : \|y - \widehat{y}\| \leq \varepsilon\}$ , and  $N_\varepsilon(\widehat{y})^c$  its complementary. Notice that

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

and the result follows.  $\square$

**Proof of Proposition 8.** From Propositions 5 and 7, we can always take sets  $\{\widehat{y}_i\}_{i=1}^{\widehat{k}} \subset \mathbb{Y}_T(\widehat{\theta}_T^{\text{wc}})$ , and  $\{y_i\}_{i=1}^k \subset \mathbb{Y}(\theta^{\text{wc}})$  in such a way that  $d_H(\{\widehat{y}_i\}_{i=1}^{\widehat{k}}, \{y_i\}_{i=1}^k) \rightarrow 0$ , and  $\widehat{k} \rightarrow k$ , almost surely (in probability), without loss of generality. As  $\{\mu_i\}_{i=1}^k$  are the Lagrange multipliers associated with the problem

$$\min_{\rho} \{\rho : Q(\theta^{\text{wc}}, y_i) \leq \rho, i = 1, \dots, k\}$$

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

$$\min_{\rho} \{\rho : Q_T(\widehat{\theta}_T^{\text{wc}}, \widehat{y}_i) \leq \rho, i = 1, \dots, \widehat{k}\},$$

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

$$L(\rho, \mu) = \rho - \sum_{i=1}^k \mu_i (Q(\theta^{\text{wc}}, y_i) - \rho), \quad L_T(\rho, \mu) = \rho - \sum_{i=1}^{\widehat{k}} \mu_i (Q_T(\widehat{\theta}_T^{\text{wc}}, \widehat{y}_i) - \rho),$$

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

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

where the  $o(1)$  term is uniform in  $\rho, \mu$ . Next,

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

when  $\max_{\theta \in \Theta, y \in \mathbb{Y}} |Q_T(\theta, y) - Q(\theta, y)| \rightarrow 0$ , and  $\widehat{\theta}_T^{\text{wc}} \rightarrow \theta^{\text{wc}}$  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  $\rho^* = Q(\theta^{\text{wc}}, y_i)$  for all  $i$ ).  $\square$

**Proof of Theorem 9.** Given an arbitrary vector  $\delta$  such that  $\delta' \delta = 1$ , let

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

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

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

uniformly in  $x$ .

By Theorem 6, there exists a positive integer  $1 \leq \widehat{k} \leq K + 1$ , vectors  $\widehat{y}_i \in \mathbb{Y}(\widehat{\theta}_T^{\text{wc}})$  and multipliers  $\widehat{\mu}_i \geq 0$  for  $i = 1, \dots, \widehat{k}$  with  $\sum_{i=1}^{\widehat{k}} \widehat{\mu}_i = 1$ , such that

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

Applying the mean value theorem,

$$0 = \sqrt{T} \sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \frac{\partial Q_T(\widehat{\theta}_T^{\text{wc}}, \widehat{y}_i)}{\partial \theta} = \sqrt{T} \sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \frac{\partial Q_T(\theta^{\text{wc}}, \widehat{y}_i)}{\partial \theta} + \sum_{i=1}^{\widehat{k}} \widehat{\mu}_i \frac{\partial^2 Q_T(\bar{\theta}_T, \widehat{y}_i)}{\partial \theta \partial \theta'} \sqrt{T} (\widehat{\theta}_T^{\text{wc}} - \theta^{\text{wc}}),$$

where  $\|\bar{\theta}_T - \theta^{\text{wc}}\| \leq \|\widehat{\theta}_T - \theta^{\text{wc}}\|$ . Under B.4, it follows that:

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

Using conditions B.2, and B.3,

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

Thus,

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

and the result follows.  $\square$

Assumption B.4 can also be established applying the following result:

**Proposition 12.** *Sufficient conditions for B.4 are*

C.1.  $B_T = \sum_{i=1}^{\hat{k}} \hat{\mu}_i \partial^2 Q_T(\theta^{\text{wc}}, \hat{y}_i) / \partial \theta \partial \theta' \rightarrow_p B$ , and

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

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

$$\left\| \frac{\partial^2 Q_T(\theta, y)}{\partial \theta \partial \theta'} - \frac{\partial^2 Q_T(\theta^{\text{wc}}, y)}{\partial \theta \partial \theta'} \right\| \leq f_T(y) \|\theta - \theta^{\text{wc}}\|^\alpha \quad (13)$$

for some  $\alpha \in (0, 1)$ , and  $E[\sup_{y \in \mathbb{Y}} |f_T(y)|] < \infty$ . For (13), it suffices that the elements in  $\partial^2 g(\theta, y) / \partial \theta \partial \theta'$  satisfy a Lipschitz condition.

**Proof of Proposition 12.** By condition C.2,

$$\begin{aligned} & E \left[ \sup_{\|\theta - \theta^{\text{wc}}\| \leq \delta} \left\| \sum_{i=1}^{\hat{k}} \hat{\mu}_i \left( \frac{\partial^2 Q_T(\theta, \hat{y}_i)}{\partial \theta \partial \theta'} - \frac{\partial^2 Q_T(\theta^{\text{wc}}, \hat{y}_i)}{\partial \theta \partial \theta'} \right) \right\| \right] \\ & \leq E \left[ \sup_{\|\theta - \theta^{\text{wc}}\| \leq \delta} \sup_{y \in \mathbb{Y}} \left\| \frac{\partial^2 Q_T(\theta, y)}{\partial \theta \partial \theta'} - \frac{\partial^2 Q_T(\theta^{\text{wc}}, y)}{\partial \theta \partial \theta'} \right\| \left\| \sum_{i=1}^{\hat{k}} \hat{\mu}_i \right\| \right] \xrightarrow{\delta \downarrow 0} 0, \end{aligned}$$

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

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

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

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  $\|\tilde{\theta}_T - \theta^{\text{wc}}\| \rightarrow_p 0$ , so we can build a sequence  $\delta_T \rightarrow 0$  such that  $\|\tilde{\theta}_T - \theta^{\text{wc}}\| \leq \delta_T$  except for sets of probability tending to zero.  $\square$

**Proof of Theorem 10.** The proof is analogous to that of Theorem 9. Assuming B.1' and  $h$  is continuous, it is satisfied that  $h_j(\hat{\theta}_T^{\text{cwc}}) = 0$  for  $j = 1, \dots, p$ , except for a set of probability tending to zero. Applying the mean value theorem to the first order necessary conditions,

$$\sum_{i=1}^{\hat{k}} \hat{\mu}_i \frac{\partial}{\partial \theta} Q_T(\hat{\theta}_T^{\text{cwc}}, \hat{y}_i) + \sum_{j=1}^p \hat{\lambda}_j \frac{\partial}{\partial \theta} h_j(\hat{\theta}_T^{\text{cwc}}) = 0, \quad h_j(\hat{\theta}_T^{\text{cwc}}) = 0, \quad j = 1, \dots, p,$$

we obtain

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

where  $H_p(\theta) = (h_1(\theta), \dots, h_p(\theta))'$ . The asymptotic normality follows analogously to Theorem 9, 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  $H = \partial H_p(\theta^{\text{wc}})/\partial \theta$ .  $\square$

**Proof of Theorem 11.** The proof is similar to that of Theorem 9. Under conditions D.1–D.4

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

applying the delta method, and for any sequence  $\tilde{\theta}_T \rightarrow_p \theta^{\text{wc}}$ ,

$$\sum_{i=1}^{\hat{k}} \hat{\mu}_i \left( T^{-1} \sum_{t=1}^T \frac{\partial}{\partial \theta} g(X_t, \hat{y}_i, \tilde{\theta}_T) \right) W_T \left( T^{-1} \sum_{t=1}^T \frac{\partial}{\partial \theta} g(X_t, \hat{y}_i, \tilde{\theta}_T) \right) \rightarrow_p B.$$

When  $\hat{\theta}_T \in \text{int}\{\Theta\}$ ,

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

Applying the mean value theorem,

$$g(X_t, \hat{y}_i, \hat{\theta}_T) = g(X_t, \hat{y}_i, \theta^{\text{wc}}) + \frac{\partial}{\partial \theta} g(X_t, \hat{y}_i, \tilde{\theta}_T) (\hat{\theta}_T - \theta^{\text{wc}}),$$

with  $\|\tilde{\theta}_T - \theta^{\text{wc}}\| \leq \|\hat{\theta}_T - \theta^{\text{wc}}\|$ , and therefore

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

It follows that:

$$\begin{aligned} B^{-1}[B + o_p(1)]\sqrt{T}(\hat{\theta}_T^{\text{wc}} - \theta^{\text{wc}}) &= -B^{-1} \sum_{i=1}^{\hat{k}} \hat{\mu}_i \left( \frac{1}{T} \sum_{t=1}^T \frac{\partial}{\partial \theta} g(X_t, \hat{y}_i, \hat{\theta}_T) \right) W_T \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T g(X_t, \hat{y}_i, \theta^{\text{wc}}) \right) \\ &\rightarrow_d N(0, B^{-1}AB^{-1}). \quad \square \end{aligned}$$

## References

Aigner, D.J.C., 1974. An appropriate econometric framework for estimating a labour-supply function from the SEO File. *Internat. Econom. Rev.* 15, 59–68.

- Balbás, A., Ibañez, A., 2002. Maxmin portfolios in models where immunization is not feasible. In: Kontoghiorghes, E., Rustem, B., Siokos, S. (Eds.), *Computational Methods in Decision Making: Economics and Finance*. Kluwer Applied Optimization Series, Boston, pp. 137–163 (Chapter 8).
- Bector, C.R., Bhatia, B.L., 1985. Sufficient optimality conditions and duality for a minimax problem. *Util. Math.* 27, 229–247.
- Berge, C., 1963. *Topological Spaces*. Macmillan, New York.
- Bickel, P.J., Wichura, M.J., 1971. Convergence criteria for multiparameter processes and some applications. *Ann. Math. Statist.* 42, 1656–1670.
- Billingsley, P., 1968. Convergence of Probability Measures. second ed. in 1999. Wiley, New York.
- Bram, J., 1966. The Lagrange multiplier theorem for max–min with several constraints. *SIAM J. Appl. Math.* 14 (4), 665–667.
- Chernozhukov, V., Hong, H., Tamer, E., 2004. Parameter set inference in a class of econometric models. Department of Economics, Northwestern University.
- Danskin, J.M., 1967. *The Theory of Max–Min*. Springer, Berlin.
- Davidson, J., 1994. *Stochastic Limit Theory*. Advanced Text in Econometrics. Oxford University Press, Oxford.
- Dem'yanov, V.F., Malozemov, V.N., 1972. *Vvedenie Minimaks*. Nauka, Moscow. (In English: *Introduction to Minimax*, Wiley, New York, 1974.).
- Devroye, L., 1987. *A Course in Density Estimation*. Progress in Probability and Statistics. Birkhäuser, Boston.
- Domiguez, M.A., Lobato, N., 2004. Consistent estimation of models defined by conditional moment restrictions. *Econometrica* 72, 1601–1615.
- Dudley, R.M., 1999. *Uniform Central Limit Theorems*. Cambridge University Press, Cambridge.
- Fishman, G.S., Rubin, D.S., 1998. Best- and worst-case variances when bounds are available for the distribution function. *Comput. Statist. Data Anal.* 36, 35–53.
- Gabler, S., 1990. *Minimax Solutions in Sampling from Finite Populations*. Lecture Notes in Statistics, vol. 64. Springer, New York.
- Goldberger, A.S., 1974. Unobservable variables in econometrics. In: Zarembka, P. (Ed.), *Frontiers in Econometrics*. Academic Press, New York.
- Griliches, Z., 1974. Errors in variables and other unobservables. *Econometrica* 42, 971–998.
- Hager, W.W., Presler, D.L., 1987. Dual techniques for minimax. *SIAM J. Control Optim.* 25, 660–685.
- 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. *J. Econom.* 30, 203–238.
- Hansen, L.P., Sargent, T.J., 2001. Robust control and model uncertainty. *Amer. Econom. Rev.* 91, 60–66.
- Hansen, L.P., Sargent, T.J., 2005. *Robustness*. Unpublished Monograph.
- Hansen, L.P., Singleton, K.J., 1982. Generalized instrumental variables estimation of nonlinear rational expectation models. *Econometrica* 50, 1269–1286.
- Hettich, R., Kortanek, K.O., 1993. Semi-infinite programming: theory, methods and applications. *SIAM Rev.* 35, 380–429.
- Honoré, B.E., Tamer, E., 2005. Bounds on parameters in panel dynamic discrete choice models. *Econometrica* 74 (3), 611–630.
- Horowitz, J.L., Manski, C.F., 1998. Censoring of outcomes and regressors due to survey nonresponse: identification and estimation using weights and imputations. *J. Econom.* 84, 37–58.
- Horowitz, J.L., Manski, C.F., 2000. Nonparametric analysis of randomized experiments with missing covariate and outcome data. *J. Amer. Statist. Assoc.* 95, 449, 77–84.
- Horowitz, J.L., Manski, C.F., 2006. Identification and estimation of statistical functionals using incomplete data. *J. Econom.* 132, 445–459.
- Horowitz, J.L., Manski, C.F., Ponomareva, M., Stoye, J., 2003. Computation of bounds on population parameters when data are incomplete. *Reliab. Comput.* 9, 419–440.
- Huber, P.J., 1964. Robust estimation of location parameter. *Ann. Math. Statist.* 37, 73–101.
- Huber, P.J., 1967. The behavior of maximum likelihood estimates under non-standard conditions. In: *Proceedings of the Fifth 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. SIAM, Philadelphia.
- Hum, D., Simpson, W., 1994. Labour supply estimation and public policy. *J. Econom. Surveys* 8, 57–81.
- Imbens, G., Manski, C.F., 2004. Confidence intervals for partially identified parameters. *Econometrica* 72, 1845–1857.
- Jennrich, R.I., 1969. Asymptotic properties of non linear least squares estimators. *Ann. Math. Statist.* 40, 633–643.
- Lehmann, E.L., 1983. *Theory of Point Estimation*. Wiley, New York.
- Lehmann, E.L., 1986. *Testing Statistical Hypotheses*. second ed. Wiley, New York.
- Lucas, R.E., 1976. Econometric policy evaluation: a critique. In: Bruckner, K., Meltzer, A. (Eds.), *The Phillips Curve and Labour Markets*, Carnegie–Rochester Conferences on Public Policy, vol. 1. North Holland, Amsterdam, pp. 19–46.
- Newey, W.K., Smith, R.J., 2004. High order properties of GMM and generalized empirical likelihood estimators. *Econometrica* 72, 219–255.
- Newey, W.K., Ramalho, J.J.S., Smith, R.J., 2005. Asymptotic bias for GMM and GEL estimators with estimated nuisance parameters. In: Rothenberg, T.J., Andrews, D.W.K., Stock, J.H. (Eds.), *Identification and Inference in Econometric Models: Essays*. Cambridge University Press, Cambridge, pp. 245–281.
- Noubiap, R.F., Seidel, W., 2001. An efficient algorithm for constructing  $\Gamma$ -minimax tests for finite parameter spaces. *Comput. Statist. Data Anal.* 36, 145–161.
- Pollard, D., 1989. Asymptotics via empirical processes. *Statist. Sci.* 4, 341–366.
- Pollard, D., 1990. *Empirical Processes: Theory and Applications*. NSF-CBMS Regional Conference Series in Probability and Statistics, vol. 2. Institute of Mathematical Statistics, Hayward, CA.
- Pötscher, B.M., Prucha, I.R., 1997. *Dynamic Nonlinear Econometric Models*. Springer, New York.
- Reemtsen, R., Görner, S., 1998. Numerical methods for semi-infinite programming: a survey. In: Reemtsen, R., Rückmann, J.J. (Eds.), *Semi-infinite Programming*. Kluwer Academic Press, Dordrecht, The Netherlands, pp. 195–275.
- Rockafellar, R.T., 1970. *Convex Analysis*. Princeton University Press, Princeton.



- Rosen, J.B., 1965. Existence and uniqueness of equilibrium points for concave  $n$ -Person games. *Econometrica* 33, 520–534.
- Rustem, B., Howe, M.A., 1997. A robust hedging algorithm. *J. Econom. Dynamics Control* 21, 1065–1092.
- Rustem, B., Howe, M.A., 2002. *Algorithms for Worst-Case Design and Applications to Risk Management*. Princeton University Press, Princeton and Oxford.
- Rustem, B., Wieland, V., Žaković, S., 2002. A continuous minimax problem and its application to inflation targeting. In: Zaccour, G. (Ed.), *Decision and Control in Management Science*. Kluwer, Dordrecht.
- Rustem, B., Wieland, V., Žaković, S., 2005. Stochastic optimization and worst-case analysis in monetary policy design. Working paper.
- Schmitendorf, W.E., 1977. Necessary conditions and sufficient conditions for static min–max problems. *J. Math. Anal. Appl.* 57, 683–693.
- Shapiro, A., 1988. First and second order optimality conditions and perturbation analysis of semi-infinite programming problems. In: Reemtsen, R., Rückmann, J.J. (Eds.), *Semi-infinite Programming*. Kluwer Academic Press, Dordrecht, The Netherlands, pp. 103–133.
- Shapiro, A., 1994. On Lipschitzian stability of optimal solutions of parametrized semiinfinite programs. *Math. Oper. Res.* 19, 743–752.
- Shimizu, K., Aiyoshi, E., 1980. Necessary conditions for min–max problems and algorithms by a relaxation procedure. *IEE Trans. Automat. Control* 25, 62–66.
- Sundaram, R.K., 1996. *A First Course in Optimization Theory*. Cambridge University Press, New York.
- Tetlow, R., von zur Muehlen, P., 2001. Robust monetary policy with misspecified models: does uncertainty always call for attenuated policy? *J. Econom. Dynamics Control* 25, 911–949.
- Tian, Y., Herzberg, A.M., 2006. A-minimax and D-minimax robust optimal designs for approximately linear Haar-wavelet models. *Comput. Statist. Data Anal.* 50 (10), 2942–2951.
- van der Vaart, A.W., Wellner, J.A., 1996. *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer, New York.
- van Geer, S., 2000. *Empirical Processes in M-Estimation*. Cambridge University Press, Cambridge.
- White, H., 1994. *Estimation, Inference, and Specification Analysis*. Cambridge University Press, New York.
- Wichura, M.J., 1969. Inequalities with applications to the weak convergence of random processes with multidimensional time parameters. *Ann. Math. Statist.* 40, 681–687.
- Wooldridge, J.M., 1994. Estimation and inference for dependent processes. In: Engle, R.F., McFadden, D.L. (Eds.), *Handbook of Econometrics*, vol. IV. North Holland, Amsterdam, pp. 2639–2738.
- Žaković, S., Rustem, B., 2003. Semi-infinite programming and applications to minimax problems. *Ann. Oper. Res.* 124, 81–110.