Improving confidence set estimation when parameters are weakly identified

Heather Battey* Qiang Feng† Richard J. Smith‡

June 17, 2016

Abstract

This paper considers inference in weakly identified moment condition models when additional partially identifying moment inequality constraints are available. The paper details the limiting distribution of the estimation criterion function exploiting both forms of moment restrictions and consequently proposes a confidence set estimator for the true parameter. The volume of the confidence set is correspondingly reduced demonstrating the benefit of exploiting moment inequality constraints in weakly identified models.

1 Introduction

This paper considers the estimation of a d_θ-vector of parameters θ_0 which is the solution to the set of moment equality restrictions

$$\mathbb{E}\left[g(Z,\theta)\right] = 0 \text{ at } \theta = \theta_0 \tag{1.1}$$

where $g(z,\theta)$ is a d_g-vector of known functions of the observation vector z and $\theta \in \Theta$ with Θ the parameter space. Estimators based on estimating equations of the form (1.1) are referred to as Z-estimators (e.g. **?**) and have found application in numerous fields, e.g., survival modelling with incomplete covariate data (**?**) and causal inference with instrumental variables (**??**).

A challenging problem arises when the identifying strength of the moment conditions (1.1) for θ_0 is weak, e.g., when instrumental variables used to construct the moment indicator $g(Z,\theta)$ are only weakly correlated with endogenous covariates. Existing inferential procedures robust to weak identification, see *inter alia* **?**, **?**, **?**, **?**, share the shortcoming that confidence set estimators for θ_0 are frequently too large to be of practical use. In many applications where weak identification is a problem, however, moment inequality conditions of the form

$$\mathbb{E}\left[m(Z,\theta)\right] \geq 0 \text{ at } \theta = \theta_0 \tag{1.2}$$

*Imperial College London and Princeton University. Email: h.battey@imperial.ac.uk.
†UIBE, Beijing. Email: q.feng@uibe.edu.cn.
‡University of Cambridge. Email: rjs27@econ.cam.ac.uk.

are often available, where $m(z,\theta)$ is a d_m-vector of functions known up to θ. This is especially so when instruments are used to overcome estimator bias induced by confounding variables, that is, when latent variables causally affect both response and covariates.

Consider the effect of smoking on health. It has been postulated (**?**) that smoking is related to health through the unobservable confounding variable, risk aversion. Thus cigarette price, being weakly correlated with cigarette consumption and uncorrelated with risk aversion, is a possible but weak instrument. Additional moment inequality information is available here since cigarette consumption and unobserved risk aversion are known to be negatively correlated. A similar scenario arises in the returns to education example of **?**, where quarter of birth is proposed as a (weak) instrument for years of schooling, and schooling and unobserved ability are known to be positively correlated, giving rise to an additional moment inequality condition.

From a technical point of view, progress is still possible in this weak instrument setting provided the strength of the correlation between the instrument and the endogenous regressor is not smaller than $\mu/\sqrt{n}$ for some $\mu \neq 0$ where n is the sample size. For this reason, data are viewed as realisations of the triangular array $\{Z_{in},\ (i=1,\ldots,n),\ (n=1,2,\ldots)\}$, and any row of the triangular array is endowed with the corresponding expectation operator $\mathbb{E}_n$, cf. Example 1 below.

Although moment inequalities taken in isolation typically only have partial or set identifying power, taken together both forms of information can result in a smaller confidence set estimator for θ_0 than that based solely on the moment equality constraints. See **?** and more recently **?** and **?** for discussions of partial identification. The concern, therefore, of this paper is the construction of a confidence set estimator for θ_0 in weakly identified models defined by (1.1) in the presence of additional partially identifying moment inequality (1.2) constraints.

To illustrate the similarities and differences between this paper and the existing literature consider the following example.

Example 1.

$$Y_i = \theta_0 X_i + \varepsilon_{1i}, \quad X_i = \gamma_{0,n} W_i + \vartheta_{0,n}\varepsilon_{1i} + \varepsilon_{2i}, \quad i = 1,\ldots,n,$$

where ε_{1i}, ε_{2i} and W_i are mutually uncorrelated. The parameter θ_0 is weakly identified if $\gamma_{0,n} = \mathbb{E}_n[X_i W_i]/\mathbb{E}_n[W_i^2] = \mu/n^{1/2}$ for $\mu \neq 0$, and $\vartheta_{0,n} = \mathbb{E}_n[X_i\varepsilon_{1i}]/\mathbb{E}_n[\varepsilon_{1i}^2] = \vartheta_0 \neq 0$, and partially identified if $\gamma_{0,n} = 0$ and $\vartheta_{0,n} = \vartheta_0 \geq 0$. **?** considers both non-weak moment equalities and moment inequalities, i.e., $\gamma_{0,n} = \gamma_0 \neq 0$ and $\vartheta_{0,n} = \vartheta_0 \geq 0$, while **?** considers $\gamma_{0,n} = \gamma_0 \neq 0$ and $\vartheta_{0,n} = c/n^{1/2} \geq 0$, where c is a constant. This paper addresses the case $\gamma_{0,n} = \mu/n^{1/2}$ and $\vartheta_{0,n} = \vartheta_0 \geq 0$.

To aid clarity, the paper focuses on the special case in which no nuisance parameters are present. For recent contributions that discuss inference in partially identified models with nuisance parameters, see **?** and **?**.

The rest of the paper is organised as follows. Section 2 defines the confidence set estimator for θ_0 and establishes its properties. Section 3 discusses its implementation with Section 4 providing an examination of the finite sample performance of the confidence set estimator.

2 Inferential procedure

Given the sample of observations $\{Z_{in},\ (i=1,\ldots,n)\}$ the interest of the paper is a nominal α-level confidence set estimator $\{\widehat{C}_n(\alpha)\}$ for θ_0 based on the continuous updating (CUE) generalized method of moments (GMM) estimation criterion (**?**); cf. **?**.

Let $\widehat{g}^n(\theta) = n^{-1}\sum_{i=1}^n g_{in}(\theta)$ and $\widehat{m}^n(\theta) = n^{-1}\sum_{i=1}^n m_{in}(\theta)$ where $g_{in}(\theta) = g(Z_{in},\theta)$ and $m_{in}(\theta) = m(Z_{in},\theta)$, $(i=1,\ldots,n)$. The CUE GMM criterion is defined as

$$\widehat{Q}_n(\theta,t) = \begin{pmatrix} \widehat{g}^n(\theta) \\ \widehat{m}^n(\theta)-t \end{pmatrix}' \widehat{V}^n(\theta)^{-1} \begin{pmatrix} \widehat{g}^n(\theta) \\ \widehat{m}^n(\theta)-t \end{pmatrix},$$

where

$$\widehat{V}^n(\theta) = n^{-1}\sum_{i=1}^n \begin{pmatrix} g_{in}(\theta)-\widehat{g}^n(\theta) \\ m_{in}(\theta)-\widehat{m}^n(\theta) \end{pmatrix} \left((g_{in}(\theta)-\widehat{g}^n(\theta))',\ (m_{in}(\theta)-\widehat{m}^n(\theta))'\right),$$

and $t \in \mathbb{R}_+^{d_m}$ is a d_m-vector of slackness parameters reflecting the inequality moment constraints (1.2). Minimisation with respect to t yields the profile CUE GMM criterion,

$$\widehat{Q}_n(\theta) = \widehat{Q}_n(\theta,\widehat{t}_n(\theta)) \quad \text{where} \quad \widehat{t}_n(\theta) = \operatorname*{arginf}_{t\in\mathbb{R}_+^{d_m}} \widehat{Q}_n(\theta,t). \tag{2.1}$$

The α-level confidence set estimator $\{\widehat{C}_n(\alpha)\}$ based on (2.1) is then defined as

$$\widehat{C}_n(\alpha) = \left\{\theta\in\Theta : n\widehat{Q}_n(\theta) \le q\right\}, \tag{2.2}$$

where q is a critical value chosen to ensure that $\lim_{n\to\infty}\mathbb{P}\mathrm{r}_n(\theta_0 \in \widehat{C}_n(\alpha)) \ge 1-\alpha$ and $\mathbb{P}\mathrm{r}_n(\cdot)$ is probability taken with respect to the joint distribution of $\{Z_{in}\}_{i=1}^n$.

Implementation of the confidence set estimator $\widehat{C}_n(\alpha)$ (2.2) requires the limit distribution of $\widehat{Q}_n(\theta_0)$. Let $\mathbb{E}_n[\cdot]$ denote expectation taken with respect to the joint distribution of $\{Z_{in}\}_{i=1}^n$. Define $g^n(\theta) = \mathbb{E}_n[\widehat{g}^n(\theta)]$ and $m^n(\theta) = \mathbb{E}_n[\widehat{m}^n(\theta)]$ likewise. The identified set is then defined by $\Theta_0 = \cap_{n=1}^\infty\{\theta\in\Theta : m^n(\theta)\ge 0\} = \{\theta\in\Theta : m(\theta)\ge 0\}$ where $m(\theta) = \lim_{n\to\infty} m^n(\theta)$; cf. **?**. The following conditions are imposed.

Condition 1 (Weak Identification)**.**

$$g^n(\theta) = k^n(\theta)/n^{1/2}, \tag{2.3}$$

where $\sup_{\theta\in\Theta}\|k^n(\theta)-k(\theta)\| = o(1)$ and $k(\theta)=0$ if and only if $\theta=\theta_0$.

Let $\rightsquigarrow$ denote weak convergence of empirical processes. Define the empirical process

$$\widehat{\Psi}_n(\theta) = n^{-1/2}\sum_{i=1}^n \left((g_{in}(\theta)) - g^n(\theta))', (m_{in}(\theta)) - m^n(\theta))'\right)'.$$

Condition 2 (Weak Convergence)**.** $\left\{\widehat{\Psi}_n(\theta) : \theta\in\Theta_0\right\} \rightsquigarrow \Psi$, where Ψ is a Gaussian process on Θ_0 with mean zero and covariance function $\Delta(\theta_1,\theta_2) = \mathbb{E}\Psi(\theta_1)\Psi(\theta_2)'$ at (θ_1,θ_2), $\theta_1,\theta_2\in\Theta_0$.

Primitive conditions for Condition 2 are given in Theorems 1.5.4 and 1.5.7 of ? requiring weak convergence of the marginals $(\widehat{\Psi}_n(\theta_1), ..., \widehat{\Psi}_n(\theta_k))$ for every finite subset $\theta_1, ..., \theta_k$ of Θ_0, stochastic equicontinuity of $\widehat{\Psi}_n(\theta)$ and total boundedness of Θ_0. For example, ? provide conditions on the data generating process that guarantee Condition 2 is satisfied. In particular, ? require $g_{in}(\theta)$ and $m_{in}(\theta)$ to be m-dependent sequences for some fixed $m \geq 0$, with the special case $m = 0$ corresponding to an independent sequence, and, in addition, to be Lipschitz continuous in expectation over Θ_0 with $2 + \epsilon$ absolute moments for some $\epsilon > 0$ uniformly over Θ_0.

Let $V(\theta) = \lim_{n\to\infty} \mathbb{V}\mathrm{ar}_n \left[n^{1/2}(\widehat{g}^n(\theta)', \widehat{m}^n(\theta)')'\right]$.

Condition 3 (Weight Matrix)**.** The weight matrix $\widehat{V}^n(\theta)$ satisfies $\sup_{\theta\in\Theta} \|\widehat{V}^n(\theta) - V(\theta)\| = o_p(1)$ where $V(\theta)$ is a non-stochastic strictly positive definite matrix uniformly over $\theta \in \Theta$. Furthermore, $\sup_{\theta\in\Theta} \|\widehat{V}^n(\theta)\| = O_p(1)$ for all $n \geq 1$.

Let b_θ denote the number of binding moment conditions at $\theta \in \Theta_0$. Also let $b_0 = b_{\theta_0}$. When the context ensures no ambiguity, the dependence of b_θ on θ is suppressed. Without loss of generality the first b_θ inequality moment conditions are assumed to be binding. Partition $\widehat{m}^n(\theta) = (\widehat{m}^n_b(\theta)', \widehat{m}^n_c(\theta)')'$ where $\widehat{m}^n_b(\theta)$ and $\widehat{m}^n_c(\theta)$ respectively correspond to the binding and non-binding inequality conditions and $c_\theta = d_m - b_\theta$. Define

$$V_b(\theta) = \lim_{n\to\infty} \mathbb{V}\mathrm{ar}_n \left[n^{1/2}(\widehat{g}^n(\theta)', \widehat{m}^n_b(\theta)')'\right].$$

Theorem 2.1. Suppose Conditions 1-3 are satisfied. Then for any constant $C > 0$,

$$\lim_{n\to\infty} \mathbb{P}\mathrm{r}_n \left\{n\widehat{Q}_n(\theta_0) > C\right\} = \sum_{j=0}^{b_0} w(b_0, b_0 - j, V_{b_0}(\theta_0))\mathbb{P}\mathrm{r}\left\{\chi^2_{d_g+j} \geq C\right\}. \tag{2.4}$$

where $w(\cdot,\cdot,\cdot)$ denotes, mutatis mutandis, the weight function defined in ? and ? and the $\chi^2_{d_g+j}$, $j = 1, \ldots, b_0$, variates are mutually independent.

We emphasise that, though superficially similar, this result is fundamentally different from that in ?. The weights $w\big(b_0, b_0 - j, V_{b_0}(\theta_0)\big)$ in (2.4) differ from those in ? and ? except in special cases because of the additional presence of the set of weak equality moment conditions (1.1). Details are contained in a supplement available upon request. An explicit expression is only available when $d_m \leq 4$; see ?. Section 3 provides details of an approximate construction for $\widehat{C}_n(\alpha)$.

If the moment inequality conditions (1.2) are omitted from $\widehat{Q}_n(\theta)$, then, for any $\theta \notin \Theta_0$, the probability that $\theta \in \widehat{C}_n(\alpha)$ is positive asymptotically whereas Proposition 3 in ? indicates this probability is 0 asymptotically when binding inequality conditions (1.2) are imposed. Moreover, if only the inequality conditions are imposed $\theta \in \widehat{C}_n(\alpha)$ w.p.a.1 for any $\theta \in \Theta_0$ such that $m(\theta) > 0$. If the weak equality conditions (1.1) are also imposed, then for $\theta \neq \theta_0$ but in Θ_0 a noncentrality parameter is present due to the presence of $k(\theta)$ which shifts the distribution of $\widehat{Q}_n(\theta)$ to the right. As a consequence there is a reduction in the probability that such a value is included in the confidence set.

3 Practicalities

Theorem 2.1 characterises the asymptotic distribution of the scaled criterion function $n\widehat{Q}_n(\theta)$ and provides the theoretical basis for the confidence set estimator. However, in practice, it is generally not possible to construct the confidence set exactly since the weights in Theorem 2.1 depend on the parameter θ_0 both directly (through $V(\theta_0)$) and indirectly (through the number b_0 of binding constraints.) Except in special cases, an explicit expression for the weights is only available when $d_m \leq 4$. To deal with the more general setting, two practical procedures to approximate the confidence set estimator are now provided.

Section 4 of **?** provides details of a numerical procedure based on **?** to obtain critical values. The resulting confidence set is conservative (**?**, Corollary 1) but is satisfactory if b_0 is both small and provides a good approximation to the maximum number of binding inequality moments (**?**). Thus, it may sometimes be beneficial to include only a small number of the most effective moment inequalities. While originally proposed for moment inequalities only, the procedure is straightforwardly extended to incorporate weak moment equalities.

The procedure due to **?** selects the set of binding moment inequalities consequently reducing the number of moment inequalities used for inference and, thus, substantially improving the empirical coverage properties of the confidence set estimator. Let

$$S\left(\begin{pmatrix} g \\ m \end{pmatrix}, V\right) = \inf_{t\geq 0} \begin{pmatrix} g \\ m-t \end{pmatrix}' V^{-1} \begin{pmatrix} g \\ m-t \end{pmatrix}.$$

Also let C denote the set of moment selection vectors $\{c\}$, whose jth element $c_j \in \{0,1\}$ corresponds to whether or not the j^{th} moment inequality is selected (see **?**). The estimated moment selection vector $\widehat{c}$ is then obtained as the solution to

$$\min_{c\in C} S\left(n^{1/2}\begin{pmatrix} \widehat{g}^n(\theta) \\ c\cdot\widehat{m}^n(\theta) \end{pmatrix}, \widehat{V}^n(\theta)\right) - |c|\sqrt{\log(n)}.$$

The d_m-dimensional vector φ is constructed according to

$$\varphi_j = \begin{cases} 0, & \text{if } \widehat{c}_j = 1, \\ \infty, & \text{if } \widehat{c}_j = 0 \end{cases}$$

The critical value is then simulated by the following steps. For each $r = 1, 2, \ldots, R$, where R is the number of replications, draw n i.i.d. copies of the random vector $Z^* \sim N(0, I_{d_g+d_m})$. Then compute $S_r = S(\widehat{\Omega}^n(\theta)^{-1/2}Z_r^* + (0_{d_g}, \varphi), \widehat{\Omega}^n(\theta))$ where 0_{d_g} is zero vector of dimension d_g,

$$\widehat{\Omega}^n(\theta) = \widehat{D}_n^{-1/2}(\theta)\widehat{V}_n(\theta)\widehat{D}_n^{-1/2} \quad \text{and} \quad \widehat{D}_n(\theta) = \operatorname{diag}(\widehat{V}^n(\theta)).$$

Finally the critical value is estimated as

$$\widehat{q}_{AS}(\theta) = \inf\left\{q : \frac{1}{R}\sum_{r=1}^{R} 1\{S_r \leq q\} \geq 1-\alpha\right\}.$$

4 Finite sample performance

All experiments concern the model

$$Y_i = \theta X_i + \varepsilon_i, \quad X_i = \frac{\pi_1}{n^{1/2}} W_{1i} + \frac{\pi_2}{n^{1/2}} W_{2i} + \nu_i, \ (i = 1, \ldots, n),$$

where π_1 and π_2 are constants and W_1 and W_2 are each uncorrelated with both error terms ε and ν which gives rise to the $d_g = 2$ weak moment equalities (cf. equation (1.1))

$$E[(Y - \theta X)W_1] = 0 \quad \text{and} \quad E[(Y - \theta X)W_2] = 0. \tag{4.1}$$

The additional variables (W_3, W_4) are each positively correlated with ε giving the $d_m = 2$ moment inequalities (cf. equation (1.2))

$$E[(Y - \theta X)W_3] \geq 0 \quad \text{and} \quad E[(Y - \theta X)W_4] \geq 0. \tag{4.2}$$

In the simulations, $\{W_1, W_2, W_3, W_4, \varepsilon, \nu\}$ are multivariate normally distributed with all covariance matrix entries zero except $\sigma_{\varepsilon\nu} = \sigma_{\varepsilon w_3} = \sigma_{\varepsilon w_4} = 0.25$, $\sigma_{\nu w_3} = 0.5$, $\sigma_{\nu w_4} = -0.5$ and $\sigma^2_{w_1} = \sigma^2_{w_2} = \sigma^2_{\varepsilon} = \sigma^2_{\nu} = \sigma^2_{w_3} = \sigma^2_{w_4} = 1$. The Supplementary Information provides an additional set of simulations in which the exogenous variables are drawn from the Student t distribution with 5 degrees of freedom yielding a very similar set of results. Thus, in this example, since the identified set $\Theta_0 = \{\theta \in \Theta : m(\theta) \geq 0\}$, Θ_0 is defined by the moment inequalities $E[(Y - \theta X)W_3] \geq 0$ and $E[(Y - \theta X)W_4] \geq 0$ and equals $[0.5, 1.5]$.

For each of 1000 Monte Carlo (MC) replications a confidence interval estimator $\widehat{C}_n(\alpha)$ is constructed for $\theta = \theta_0 = 1$ based on the **?** procedure using both the weak moment equality (cf. equation (4.1)) and the partial moment inequality (cf. equation (4.2)) information. **?** also adopts this procedure but using the partial information alone. To distinguish these approaches the notation R(P) is adopted for the **?** procedure and R(W&P) for the extension of **?** incorporating the additional weak moment equalities information. Confidence set construction based on the **?** procedure is also considered, adapted to exploit both weak and partial information, and is denotes by AS(W&P). R(W&P) and AS(W&P) are also compared to constructions based on **?** (SW(W)) and **?** (K(W)) based on the weak moment equalities alone.

Table 1 compares the mean width (over MC replications) of the various confidence interval estimators around $\theta_0 = 1$ with $\pi_1 = 1$ and $\pi_2 = 2$. The third and fourth columns, R(W&P) and AS(W&P), both use the weak, (4.1), and partial, (4.2), identifying information. Their differences arise because of the different implementations discussed above. The difference between the third and last columns is due solely to the additional weakly identifying information used in the construction R(W&P) over R(P), the imposition of the additional inequality conditions reducing the width of the confidence intervals appreciably.

Table 2 presents the coverage properties of the various estimators, confirming that those based on our approach, R(W&P) and AS(W&P), are conservative. Note that despite their greater length, the estimators SW(W)and K(W) based solely on weakly identifying moment conditions have lower coverage.

n	$1-\alpha$	R(W&P)	AS(W&P)	SW(W)	K(W)	R (P)
100	0.9	1.943	2.041	2.526	2.770	2.226
500	0.9	1.257	1.192	2.442	2.582	1.358
1000	0.9	1.146	1.093	2.496	2.524	1.205
100	0.95	2.229	2.401	2.979	3.107	2.469
500	0.95	1.372	1.316	2.934	3.036	1.424
1000	0.95	1.217	1.167	2.955	3.030	1.251
100	0.99	2.821	3.154	3.488	3.608	3.023
500	0.99	1.590	1.543	3.488	3.553	1.602
1000	0.99	1.351	1.317	3.549	3.621	1.361

Table 1: Mean width of confidence interval around $\theta_0 = 1$: 1000 MC replications and n observations.

n	$1-\alpha$	R(W&P)	AS(W&P)	SW(W)	K(W)	R(P)
100	0.9	0.971	0.971	0.830	0.905	1.000
500	0.9	0.976	0.969	0.895	0.905	1.000
1000	0.9	0.974	0.967	0.916	0.914	1.000
100	0.95	0.985	0.993	0.934	0.943	1.000
500	0.95	0.985	0.981	0.946	0.951	1.000
1000	0.95	0.984	0.978	0.942	0.957	1.000
100	0.99	0.997	0.998	0.983	0.982	1.000
500	0.99	0.998	0.999	0.986	0.987	1.000
1000	0.99	0.998	0.997	0.988	0.991	1.000

Table 2: Coverage probabilities based on 1000 MC replications, n observations and $H_0 : \theta = \theta_0 = 1$.

Figure 1 plots the empirical power curves for various combinations of π_1 and π_2; *viz.* $\{\pi_1 = 1, \pi_2 = 1\}$, $\{\pi_1 = 1, \pi_2 = 2\}$, $\{\pi_1 = 2, \pi_2 = 2\}$ and $\{\pi_1 = 2, \pi_2 = 3\}$. In all cases $\theta = \theta_0 = 1$ and the sample size is $n = 1000$. The SW(W) and K(W) empirical power curves are much flatter than those of the AS(W&P), R(W&P) and R(P) procedures. Although SW(W) and K(W) have somewhat higher power inside the identified region, the power of AS(W&P), R(W&P) and R(P) is far greater outside the identified region. Since the moment equalities constitute only weakly identifying information, the power of SW(W) and K(W) is expected to be relatively low across the whole parameter space. In contrast, the information provided by the moment inequality constraints is partially identifying and, thus, powerful for detecting a false null hypothesis since the true $\theta = \theta_1$ then lies outside the identified set, thereby violating the inequalities. Inside the identified region, the moment inequalities are uninformative and, thus, their inclusion effectively adds noise, explaining the lower power of AS(W&P) and R(W&P) as compared with that of SW(W) and K(W). The AS(W&P) and R(W&P) approaches dominate R(P), emphasising the value of the weakly identifying moment information. In particular, as the values of π_1 and π_2 increase, the AS(W&P) and R(W&P) power curves lie significantly above that of R(P).

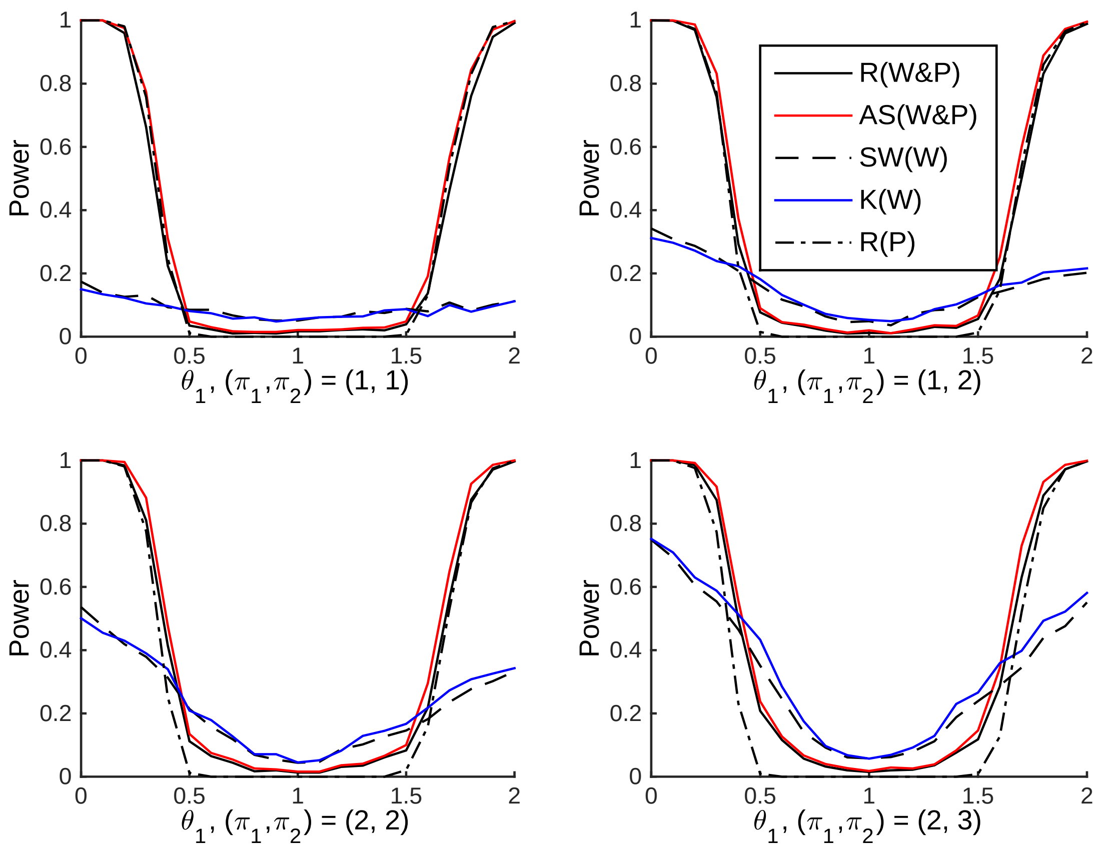

Figure 1: Estimated power against the alternative $\theta = \theta_1 \in [0, 2]$ for $\theta_0 = 1$, $n = 1000$ and various values of π_1 and π_2.

5 Proof of Theorem 2.1

Let

$$\bar{v}^n(\theta) = n^{1/2}\begin{pmatrix} \widehat{g}^n(\theta) - g^n(\theta) \\ \widehat{m}^n(\theta) - m^n(\theta) \end{pmatrix}, \quad \bar{v}_b^n(\theta) = n^{1/2}\begin{pmatrix} \widehat{g}^n(\theta) - g^n(\theta) \\ \widehat{m}_b^n(\theta) - m_b^n(\theta) \end{pmatrix}.$$

where $m^n(\theta) = (m_b^n(\theta)', m_c^n(\theta)')'$ is partitioned conformably with $\widehat{m}^n(\theta)$. By Condition 1,

$$\begin{aligned} n\widehat{Q}_n(\theta) &= \min_{t\geq 0}\left(\bar{v}^n(\theta) + \begin{pmatrix} n^{1/2}g^n(\theta) \\ n^{1/2}m^n(\theta) - t \end{pmatrix}\right)' \widehat{V}^n(\theta)^{-1}\left(\bar{v}^n(\theta) + \begin{pmatrix} n^{1/2}g^n(\theta) \\ n^{1/2}m^n(\theta) - t \end{pmatrix}\right) \\ &= \min_{s\geq -n^{1/2}m^n(\theta)}\left(\bar{v}^n(\theta) - \begin{pmatrix} -k^n(\theta) \\ s \end{pmatrix}\right)' \widehat{V}^n(\theta)^{-1}\left(\bar{v}^n(\theta) - \begin{pmatrix} -k^n(\theta) \\ s \end{pmatrix}\right) \\ &= \min_{s_b\geq 0,\, s_c\geq -n^{1/2}m_c^n(\theta)}\left(\bar{v}^n(\theta) - \begin{pmatrix} -k^n(\theta) \\ s \end{pmatrix}\right)' \widehat{V}^n(\theta)^{-1}\left(\bar{v}^n(\theta) - \begin{pmatrix} -k^n(\theta) \\ s \end{pmatrix}\right), \end{aligned}$$

where $s = (s_b', s_c')'$ is partitioned conformably with $m^n(\theta)$. Since $m_c^n(\theta) > 0$, $n^{1/2}m_c^n(\theta) \to \infty$ and thus, invoking also Condition 3,

$$n\widehat{Q}_n(\theta) - \min_{s_b\geq 0, s_c\in\ \mathbb{R}^c}\left(\bar{v}^n(\theta) - \begin{pmatrix} -k(\theta) \\ s \end{pmatrix}\right)' V(\theta)^{-1}\left(\bar{v}^n(\theta) - \begin{pmatrix} -k(\theta) \\ s \end{pmatrix}\right) = o_p(1). \tag{5.1}$$

By Lemma 1 of **?**

$$\begin{aligned} &\min_{s_b\in\mathbb{R}_+^b, s_c\in\mathbb{R}^c}\left(\bar{v}^n(\theta) - \begin{pmatrix} -k(\theta) \\ s \end{pmatrix}\right)' V(\theta)^{-1}\left(\bar{v}_n(\theta) - \begin{pmatrix} -k(\theta) \\ s \end{pmatrix}\right) \\ &= \min_{s_b\in\mathbb{R}_+^b}\left(\bar{v}_b^n(\theta) - \begin{pmatrix} -k(\theta) \\ s_b \end{pmatrix}\right)' V_b(\theta)^{-1}\left(\bar{v}_b^n(\theta) - \begin{pmatrix} -k(\theta) \\ s_b \end{pmatrix}\right). \end{aligned}$$

By Condition 2, $\bar{v}_b^n \rightsquigarrow \bar{v}_b$ where $\bar{v}_b$ is a zero mean Gaussian process on Θ_0, thus, for all $\theta \in \Theta_0$ $\bar{v}_b(\theta) \sim N(0, V_b(\theta))$. It follows that, for any $\theta \in \Theta_0$,

$$n\widehat{Q}_n(\theta) \to_d \min_{s_b\in\mathbb{R}_+^b}\left(\bar{v}_b - \begin{pmatrix} -k(\theta) \\ s_b \end{pmatrix}\right)' V_b(\theta)^{-1}\left(\bar{v}_b - \begin{pmatrix} -k(\theta) \\ s_b \end{pmatrix}\right) = Q^b(\theta),$$

and using the results of **?** based on **?**,

$$\mathbb{P}\mathrm{r}\left\{Q^b(\theta_0) \geq C\right\} = \sum_{j=0}^{b} w(b, b-j, V_b(\theta_0))\mathbb{P}\mathrm{r}\left\{\chi^2_{d_g+j} \geq C\right\}.$$

□

Acknowledgements

Feng thanks the National Science Foundation of China (grant 71301026) for financial support.

Supplementary material to

Improving confidence set estimation when parameters are weakly identified

Heather Battey* Qiang Feng† Richard J. Smith‡

Abstract

This document contains the supplementary material to the paper "Improving confidence set estimation when parameters are weakly identified". In Appendix A we provide additional simulations.

A Additional simulations

The data are generated in the same way as in Section 4 but with the exogenous variables W_1 and W_2 each drawn independently from a Student t distribution with 5 degrees of freedom. The results are very similar to those appearing in Section 4. We only present the results for the $\alpha = 0.05$ case.

n	$1-\alpha$	R(W&P)	AS(W&P)	SW(W)	K(W)	R(P)
100	0.95	2.269	2.552	2.903	3.135	2.586
500	0.95	1.444	1.421	2.791	2.920	1.517
1000	0.95	1.305	1.278	2.763	2.879	1.351

Table 3: Mean width of confidence interval around $\theta_0 = 1$: 1000 MC replications and n observations.

n	$1-\alpha$	R(W&P)	AS(W&P)	SW(W)	K(W)	R(P)
100	0.95	0.993	0.996	0.955	0.953	1.000
500	0.95	0.993	0.989	0.952	0.954	1.000
1000	0.95	0.984	0.982	0.944	0.945	1.000

Table 4: Coverage probabilities based on 1000 MC replications, n observations and $H_0 : \theta = \theta_0 = 1$ when exogenous variables are generated from a Student t distribution with 5 degrees of freedom

*Department of Mathematics, Imperial College London, 180 Queen's Gate, London, SW7 2AZ. Email: h.battey@imperial.ac.uk and Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08540. Email: hbattey@princeton.edu.

†1223 Boxue Building, UIBE, Beijing, 100029, China. Email: q.feng@uibe.edu.cn.

‡Faculty of Economics, University of Cambridge, Sidgwick Avenue, Cambridge, CB3 9DD. Email: rjs27@cam.ac.uk.

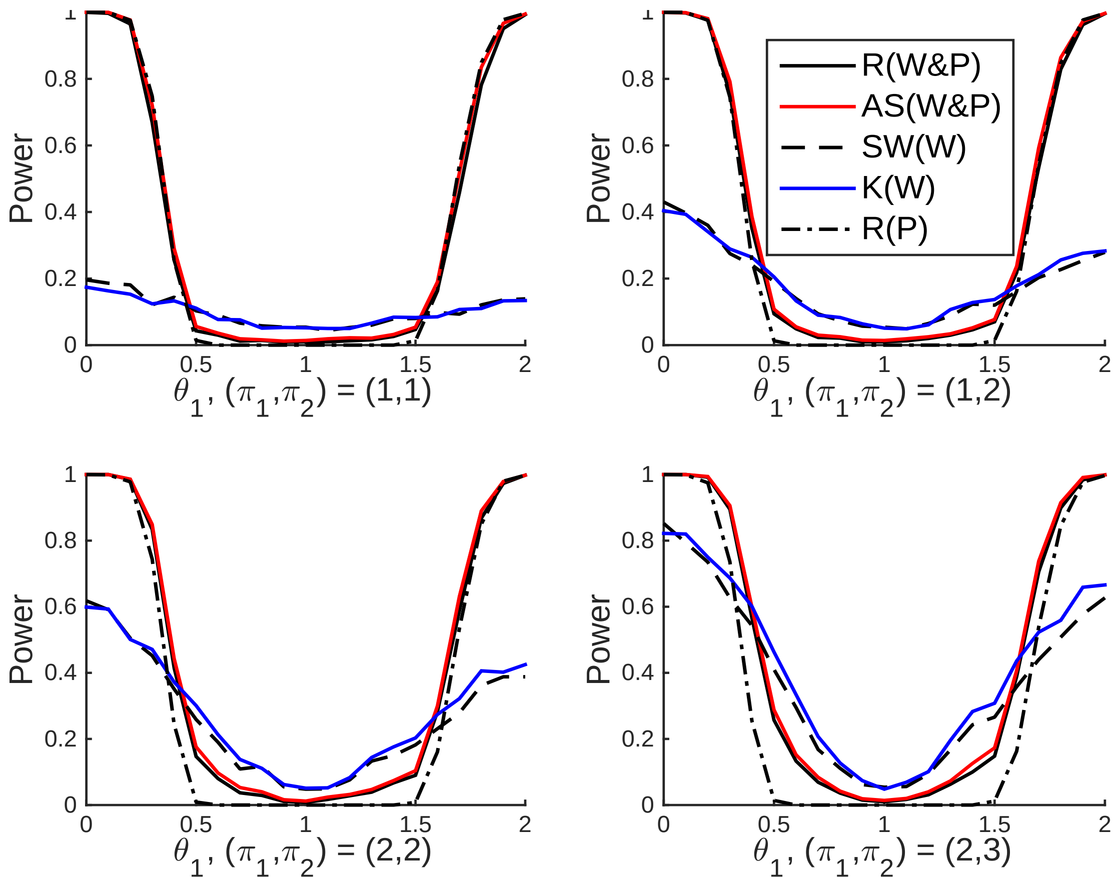

Figure 2: Estimated power against the alternative $\theta = \theta_1$ (ranging from 0 to 2) for $\theta_0 = 1$, $n = 1000$ and various values of π_1 and π_2. Exogenous variables generated from a t distribution with 5 degrees of freedom.