

# Weak identification robust tests for subvectors using implied probabilities\*

Marine Carrasco<sup>†</sup> and Saraswata Chaudhuri<sup>‡</sup>

January 2025

## Abstract

This paper develops tests for hypotheses concerning subvectors of parameters in models defined by moment conditions. It is well-known that conventional tests such as Wald, Likelihood-ratio and Score tests tend to overreject when the identification is weak. To prevent uncontrolled size distortion and introduce refined finite-sample performance, we extend the projection based test of Chaudhuri and Zivot (2011) to a modified version of the score test using implied probabilities obtained by Information theoretic criteria. Our test is performed in two steps, the first step reduces the space of parameter candidates, while the second one involves the modified score test mentioned earlier. We derive the asymptotic properties of this procedure for the entire class of Generalized Empirical Likelihood implied-probabilities. Simulations show that the test has very good finite-sample size and power. Finally, we apply our approach to the veteran earnings and find a negative impact of veteran status.

Keywords: empirical likelihood, entropy, exponential tilting, projection based test, score test, weak instruments.

---

\*The authors thank Eric Renault for his helpful comments.

<sup>†</sup>Department of Economics, University of Montreal, Citeq, Montreal. Email: marine.carrasco@umontreal.ca. Carrasco gratefully acknowledges partial funding from FRQSC and SSHRC.

<sup>‡</sup>Department of Economics, McGill University & Citeq, Montreal. Email: saraswata.chaudhuri@mcgill.ca.

# 1 Introduction

We are interested in developing tests for hypothesis concerning subvectors of an unknown parameter  $\theta \in \mathbb{R}^{d_\theta}$ . The true value of the parameter  $\theta$ , denoted  $\theta^0$ , satisfies a vector of moment conditions:

$$E[g(W; \theta^0)] = 0$$

where the vector  $g \in \mathbb{R}^{d_g}$  is known and  $d_g \geq d_\theta$ .

Based on a random sample  $W_i, i = 1, 2, \dots, n$ , the standard approach of inference is to conduct a Wald test based on the Generalized Method of Moments (GMM) estimator of  $\theta$  or a score test. Wald tests have been shown to be inappropriate in the presence of weak identification (Dufour (1997)). Moreover, the GMM-based score test proposed by Newey and West (1987) is plagued by size distortions under common scenarios such as skewed moment vectors or models with weak identification. To improve the finite sample properties of this test, Chaudhuri and Renault (2015) and Chaudhuri and Renault (2020) propose to replace the uniform weights by implied probabilities obtained from an Information Theory criterion. These probabilities exploit the information from the model, namely that  $E[g(W_i, \theta)] = 0$ . So the implied probabilities  $\hat{\pi}_i$  are selected such that the moments hold exactly:

$$\sum_{i=1}^n \hat{\pi}_i g(W_i; \theta) = 0.$$

However, given that the number of moments,  $d_g$ , is smaller than the sample size  $n$ , there is an infinity of possibilities for  $\hat{\pi}_i, i = 1, 2, \dots, n$ . The estimation of  $\pi_i$  is an ill-posed problem. Which distribution should be used? A solution inspired from the entropy literature is to select the distribution obtained by minimizing the Cressie-Read divergence measure under the moment restrictions. Equivalently, one could also work with the Generalized Empirical Likelihood (GEL) that is characterized by the dual problem of this Cressie-Read divergence minimization. Two notable members of this class are the Empirical Likelihood estimator and the Exponential Tilting estimator (see Newey and Smith (2004)). All these estimators can be viewed as Information Theory estimators (see Kitamura and Stutzer (1997) and Golan (2006)).

Chaudhuri and Renault (2020) focus on tests for the entire parameter vector, i.e.  $H_0 : \theta = \theta_0$  and they show that implied probability based score tests lead to improved finite sample properties compared to the conventional score test. In particular, they have better size control and remain powerful.

In this paper, we are concerned with testing subsets of parameters  $\theta_1 \subset \theta$ . More precisely,

we want to test  $H_0 : \theta_1 = \theta_{10}$ . The subset version of score tests suffers from important size distortion as shown by [Guggenberger et al. \(2012\)](#). To address this issue, we suggest to use the projection-based test developed by [Chaudhuri and Zivot \(2011\)](#) coupled with the score test that includes the GEL implied probabilities.

The contribution of our paper is to provide a framework that opens up the possibility of applying any type of the Generalized Empirical Likelihood or Cressie-Read implied probabilities to the type of score tests discussed in [Chaudhuri and Zivot \(2011\)](#); see also [Smith \(1997\)](#), [Newey and Smith \(2004\)](#). We derive the asymptotic properties of the resulting tests using the properties of the implied probabilities obtained in [Chaudhuri and Renault \(2020\)](#) and generalized to include all the GEL estimators. Special care is taken to allow for weak identification. The simulations show that these tests perform well in terms of finite-sample size and exhibits strong power under the alternative. We complete the paper with an empirical illustration examining the effect of veteran status on earnings. Using our proposed test, we construct confidence intervals for the returns to veteran status on earnings, leveraging instrumental variables. This analysis, inspired by [Chaudhuri and Rose \(2009\)](#), builds on the seminal natural experiment framework developed by [Angrist \(1990\)](#) and [Card \(1995\)](#), which earned them the 2021 Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel. Our findings provide evidence of a negative impact of veteran status on earnings.

The related literature is vast. The application of Information Theory measures to the estimation of econometric models goes back to [Golan et al. \(1996\)](#), [Kitamura and Stutzer \(1997\)](#), [Imbens et al. \(1998\)](#), etc. The use of the implied probabilities in the context of testing of hypothesis in the GMM setup was pioneered by [Guggenberger and Smith \(2005\)](#) and further developed by [Caner \(2010\)](#), [Chaudhuri and Renault \(2015\)](#), and [Chaudhuri and Renault \(2020\)](#). Our current paper builds on this literature. Extensive research in econometrics has demonstrated that testing of subsets of parameters in the face of commonly encountered problems such as weak identification is a much more difficult problem than test of the full parameter vector studied in [Chaudhuri and Renault \(2020\)](#); see, e.g., [Guggenberger et al. \(2012\)](#), [Andrews et al. \(2019\)](#). [Chaudhuri and Zivot \(2011\)](#) provide an early contribution to weak identification robust testing of subsets of parameters that was subsequently extended and refined in [Andrews \(2017\)](#) and [Andrews \(2018\)](#), and that is particularly suitable for application of the Information Theory. In the current paper, we demonstrate that application of Information Theory in the form of the implied probabilities to [Chaudhuri and Zivot \(2011\)](#)'s test for subsets of parameters delivers improved finite-sample performance.

The remainder of this paper is organized as follows. Section 2 describes the GMM framework and the implied probabilities in the context of the null hypothesis for subsets of parameters that is the focus of our interest. Section 3 discusses the score test for subsets of parameters and establishes its asymptotic properties. Section 4 provides evidence of the improved performance of this test using simulation results in empirically relevant settings. Section 5 includes the empirical application. Finally, Section 6 concludes. The main proofs are collected in Appendix.

## 2 Implied probabilities for hypothesis on subsets of parameters in the GMM framework

### 2.1 Background

Let  $W_1, \dots, W_n$  be independent and identically distributed (i.i.d.) copies of the random variable  $W$ . Let  $g(W; \theta) : \mathbb{R}^{d_w} \times \Theta \mapsto \mathbb{R}^{d_g}$  be the  $d_g$ -dimensional moment vector,  $W$  a  $d_w$ -dimensional random vector,  $\Theta \subseteq \mathbb{R}^{d_\theta}$  the parameter space and let  $d_g \geq d_\theta$ . Suppose that we have a set of moment restrictions:

$$E[g(W; \theta^0)] = 0 \tag{1}$$

which holds for the true value of the parameter  $\theta^0$ .

Our goal is testing of hypotheses on subsets of parameters, i.e., a subvector of  $\theta$ . Without loss of generality, let  $\theta = (\theta'_1, \theta'_2)'$ , and let the null hypothesis of interest be

$$H_0 : \theta_1 = \theta_{10}. \tag{2}$$

The parameter  $\theta_1$  is the parameter of primary interest while  $\theta_2$  is the nuisance parameter.

The usual approach to tackle this problem consists in estimating  $\theta$  by constrained GMM. Constrained estimators are obtained by imposing the null hypothesis, the estimator takes the form  $\theta = (\theta'_{10}, \theta'_2)'$  that restricts  $\theta_1$  by  $H_0$  but lets the  $\theta_2$  parameters be unrestricted. Given a first-step consistent estimator of  $\theta$ , denoted  $\bar{\theta}$ , the constrained GMM estimator is solution of

$$\tilde{\theta} = \arg \min_{\theta \in \Theta_1} Q_n(\theta) \equiv \bar{g}_n(\theta)' \hat{\Omega}_n(\bar{\theta})^{-1} \bar{g}_n(\theta)$$

where  $\Theta_1$  is the set of elements  $(\theta'_1, \theta'_2)'$  of  $\Theta$  such that  $\theta_1 = \theta_{10}$ ,  $\bar{g}_n(\theta) := \sum_{i=1}^n g(W_i; \theta)/n$ , and  $\hat{\Omega}_n(\theta) := \frac{1}{n} \sum_{i=1}^n g(W_i; \theta)g'(W_i; \theta)$ . Let  $\bar{G}_n(\theta) = (1/n) \sum_{i=1}^n \partial g(W_i; \theta)/\partial \theta'$ . The score-test

proposed by [Newey and West \(1987\)](#) is

$$LM_n = \left( \frac{\partial Q_n(\tilde{\theta})}{\partial \theta} \right)' \mathcal{I}_n^{-1}(\tilde{\theta}) \left( \frac{\partial Q_n(\tilde{\theta})}{\partial \theta} \right)$$

where

$$\begin{aligned} \partial Q_n(\theta) / \partial \theta &= \bar{G}_n(\theta)' \hat{\Omega}_n(\theta)^{-1} \bar{g}_n(\theta) \text{ and} \\ \mathcal{I}_n(\theta) &= \bar{G}_n(\theta)' \hat{\Omega}_n(\theta)^{-1} \bar{G}_n(\theta). \end{aligned} \quad (3)$$

According to [Chaudhuri and Renault \(2015\)](#) and [Chaudhuri and Renault \(2020\)](#), the poor finite sample properties of the score test can be improved by replacing the averages in  $\bar{G}_n(\theta)$  and  $\hat{\Omega}_n(\theta)$  by weighted sum using Information Theory. Instead of averaging using an equal weight  $1/n$ , one should use the implied probabilities obtained from Information theoretic criteria. The criterion considered is the Cressie-Read family.

The optimization problem solved by the implied probabilities  $\hat{\pi}_n^{(\gamma)}(\theta)$  for  $\theta \in \Theta$  is

$$\min_{\pi \in \mathbb{R}^n} \frac{1}{\gamma(\gamma+1)} \sum_{i=1}^n [(n\pi_i)^{1+\gamma} - 1] \quad \text{subject to} \quad \sum_{i=1}^n \pi_i = 1 \text{ and } \sum_{i=1}^n \pi_i g(W_i; \theta) = 0. \quad (4)$$

The objective function (4) is defined for any real  $\gamma$ , including the two limit cases  $\gamma \rightarrow 0$  and  $\gamma \rightarrow -1$ .  $\gamma = -1$  corresponds to Empirical likelihood (EL), and  $\gamma = 1$  corresponds to the Euclidean empirical likelihood (EEL), and  $\gamma \rightarrow 0$  to the Kullback Leibler Information Criterion (KLIC) that is consistent with Shannon's entropy. The so-called Generalized empirical likelihood (GEL) estimator of  $\theta$  is obtained by minimizing the criterion in (4) with respect of  $\theta$  or alternatively by minimizing the dual problem based on the Lagrange multipliers associated with the constraints  $\sum_{i=1}^n \pi_i g(W_i; \theta) = 0$ , see [Guggenberger and Smith \(2005\)](#) and [Chaudhuri and Renault \(2020\)](#). Here, however, the aim is not to estimate  $\theta$  but to perform a test. Therefore, we need to go further than the aforementioned references and devise the two-step approach described in Section 3 to address the additional inferential issues of uncontrolled over-rejection of the truth without unnecessary loss in power.

## 2.2 Assumptions

Let us define

$$V(\theta) := \text{Var}(\sqrt{n}\hat{g}_n(\theta)) \quad \text{and} \quad \hat{V}_n(\theta) := \frac{1}{n} \sum_{i=1}^n g(W_i; \theta) [g(W_i; \theta) - \bar{g}_n(\theta)]'.$$

Remark that this definition of  $\hat{V}_n(\theta)$  corresponds to the appropriate estimator of  $V(\theta)$  for EEL estimator.

Consider a sequence of subsets  $\Theta_n : n \geq 1$  of  $\Theta$  containing  $\theta^0$ .  $\Theta_n$  is a neighborhood of  $\theta^0$  which width depends on the identification strength in (1). Typically,  $\Theta_n$  is narrower for strongly identified parameters and wider for weakly identified parameters.

For illustration, consider the example of the linear instrumental variable regression:

$$y_i = X_i \theta^0 + u_i,$$

where the endogenous variable  $X_i$  is a scalar random variable related to the instruments  $Z_i$  through the reduced-form equation

$$X_i = Z_i' \Pi + V_i$$

where  $E[Z_i u_i] = 0$ . Let  $W_i = (y_i, X_i, Z_i)'$ , the moment condition corresponding to the orthogonality between  $Z_i$  and  $u_i$  is

$$g(W_i; \theta) = Z_i (y_i - X_i \theta).$$

If  $\Pi$  is non null and independent of  $n$ , the instruments  $Z_i$  are strongly correlated with the endogenous regressor  $X_i$  and hence the instruments are said to be strong. In that case,  $\theta$  is identified in the sense that  $Eg(W_i; \theta) = 0 \Leftrightarrow \theta = \theta^0$ . Then, the GMM estimator is consistent with  $\sqrt{n}$  rate of convergence. When  $\Pi = C/\sqrt{n}$ , the correlations between  $Z_i$  and  $X_i$  go to zero, and the instruments are said to be weak (in the sense of [Staiger and Stock \(1997\)](#)). In that case, the GMM estimator of  $\theta$  is not consistent because  $E[g(W_i; \theta)] \rightarrow 0$  as  $n \rightarrow \infty$  for all  $\theta \neq \theta^0$  (with  $\theta - \theta^0$  fixed). Then, the standard confidence intervals and tests are not reliable. In the semi-weak/semi-strong case, i.e. when  $\Pi = C/n^\phi$  with  $0 < \phi < 1/2$ , the GMM estimator of  $\theta$  is consistent with a slower rate of convergence than the usual  $\sqrt{n}$ .

We will maintain Assumptions 1 and 2 below to show the asymptotic equivalence of the implied probabilities for  $\theta \in \Theta_n$ ; see [Guggenberger and Smith \(2005\)](#) and [Chaudhuri and Renault \(2020\)](#) for more discussion.

**Assumption 1:**

- (i)  $\sup_{\theta \in \Theta_n} \|E[\bar{g}_n(\theta)]\| = O\left(\frac{1}{\sqrt{n}}\right)$ .
- (ii)  $\max_{1 \leq i \leq n} \sup_{\theta \in \Theta_n} \|g(W_i; \theta)\| = o_p(\sqrt{n})$ .
- (iii)  $\sup_{\theta \in \Theta_n} \|g(W_i; \theta)\| = O_p(1)$  for  $i = 1, \dots, n$ .
- (iv)  $\sup_{\theta \in \Theta_n} \|\bar{g}_n(\theta) - E[\bar{g}_n(\theta)]\| = O_p\left(\frac{1}{\sqrt{n}}\right)$ .

**Assumption 2:**

- (i)  $\sup_{\theta \in \Theta_n} \|\hat{\Omega}_n(\theta) - V(\theta)\| = o_p(1)$ ,  $\sup_{\theta \in \Theta_n} \|\hat{V}_n(\theta) - V(\theta)\| = o_p(1)$ ,  
 $\sup_{\theta \in \Theta_n} \|\hat{\Omega}_n^{-1}(\theta) - V^{-1}(\theta)\| = o_p(1)$  and  $\sup_{\theta \in \Theta_n} \|\hat{V}_n^{-1}(\theta) - V^{-1}(\theta)\| = o_p(1)$ .
- (ii)  $0 < \inf_{\theta \in \Theta_n} b_{\min}(\theta) < \sup_{\theta \in \Theta_n} b_{\max}(\theta) < +\infty$  where  $b_{\min}(\theta)$  and  $b_{\max}(\theta)$  stand for the smallest and largest eigenvalues respectively of  $V(\theta)$ .

Assumption 1 is not restrictive if  $\Theta_n$  is reduced to  $\theta^0$  for all  $n$ . In that case, Assumptions 1(i) and 1(iii) are fulfilled by definition. Assumptions 1(ii) and 1(iv) follow from the fact that  $g(W_i; \theta^0)$  is i.i.d. with zero mean and finite variance. Assumption 1(ii) is a consequence of the Borel-Cantelli Lemma. Assumption 1(iv) can be proved by Lindeberg-Levy Central Limit Theorem. Validity of Assumption 1 when  $\Theta_n$  is a local neighborhood of  $\theta^0$ , where the definition of local depends on the identification strength of  $\theta$ , follows under mild conditions.

Regarding Assumption 2(i), it requires the uniform law of large numbers for the sample covariance matrix only. The two convergences on the first line of (i) are equivalent, provided Assumptions 1(i) and 1(iv) hold. The same is true for the convergence on the second line under the extra condition of Assumption 2(ii), which ensures that the population covariance matrix is positive definite and finite.

### 2.3 Properties of GEL implied probabilities

In this section, we investigate the properties of weighted sums based on implied probabilities. To do so, it is convenient to use the dual representation of the estimators introduced in Section 2.1.

The GEL class of estimators of  $\theta^0$  is indexed by the function  $\rho$  and is defined as

$$\hat{\theta}_{\rho,n} := \arg \min_{\theta \in \Theta} \sup_{\lambda \in \Lambda_n(\theta)} \hat{Q}_{\rho,n}(\theta, \lambda)$$

where 
$$\hat{Q}_{\rho,n}(\theta, \lambda) := \frac{1}{n} \sum_{i=1}^n \rho(\lambda' g(W_i; \theta)) - \rho(0),$$

and 
$$\Lambda_n(\theta) := \{\lambda \in \mathbb{R}^k : \lambda' g(W_i; \theta) \in \mathcal{O}, \forall i = 1, \dots, n\}.$$

Different choices of  $\rho(\cdot)$  lead to different GEL estimators. The Continuous-Updating GMM or Euclidean empirical likelihood (EEL) estimator is a special case with  $(\rho(v) = -(1+v)^2/2, \mathcal{O} = \mathbb{R})$  corresponding to  $\gamma \rightarrow -1$  in Equation (4), the empirical likelihood (EL) estimator  $(\rho(v) = \ln(1-v), \mathcal{O} = (-1, \infty))$  corresponds to  $\gamma = 1$ , the exponential tilting (ET) estimator  $(\rho(v) = -\exp(v), \mathcal{O} = \mathbb{R})$  to  $\gamma \rightarrow 0$ , etc. all of which satisfy Assumption  $\rho$  below.

**Assumption  $\rho$ :** (GEL function)

$\rho : \mathcal{O} \mapsto \mathbb{R}$  is a continuous function such that

- (i)  $\rho$  is concave on its domain  $\mathcal{O}$  which is an open interval containing 0.
- (ii)  $\rho$  is twice continuously differentiable on its domain. Defining  $\rho_r(v) := \partial^r \rho(v)/\partial v^r$  for  $r = 1, 2$  and  $\rho_r := \rho_r(0)$ , let  $\rho_1 = \rho_2 = -1$  (standardization for convenience).
- (iii) There exists a positive constant  $b$  such that for each  $v \in \mathcal{O}$ ,  $|\rho_2(v) - \rho_2(0)| \leq b \times |v|$  hold.

The desirable higher-order properties of the GEL estimators are due to the GEL first order condition which, assuming differentiability of the moment vector  $g(w; \theta)$  with respect to  $\theta$ , is given by

$$\left[ \sum_{i=1}^n \pi_{\rho, i, n}(\hat{\theta}_{\rho, n}) G_i(\hat{\theta}_{\rho, n}) \right]' \left[ \sum_{i=1}^n \kappa_{\rho, i, n}(\hat{\theta}_{\rho, n}) g(W_i; \hat{\theta}_{\rho, n}) g'(W_i; \hat{\theta}_{\rho, n}) \right]^{-1} \bar{g}_n(\hat{\theta}_{\rho, n}) = o_P \left( \frac{1}{\sqrt{n}} \right)$$

where for given  $\theta$  and  $\rho(\cdot)$ ,  $\bar{g}_n(\theta) := \frac{1}{n} \sum_{i=1}^n g(W_i; \theta)$ ,  $G_i(\theta) := \frac{\partial}{\partial \theta'} g(W_i; \theta)$ ,

$$\lambda_{\rho, n}(\theta) \quad : \quad = \arg \sup_{\lambda \in \Lambda_n(\theta)} \hat{Q}_{\rho, n}(\theta, \lambda), \quad (5)$$

$$\pi_{\rho, i, n}(\theta) \quad : \quad = \frac{\rho_1(\lambda'_{\rho, n}(\theta) g(W_i; \theta))}{\sum_{j=1}^n \rho_1(\lambda'_{\rho, n}(\theta) g(W_j; \theta))} \quad : \quad \text{implied probabilities from GEL}, \quad (6)$$

$$\kappa_{\rho, i, n}(\theta) \quad : \quad = \frac{\kappa_{\rho}(\lambda'_{\rho, n}(\theta) g(W_i; \theta))}{\sum_{j=1}^n \kappa_{\rho}(\lambda'_{\rho, n}(\theta) g(W_j; \theta))}, \quad \kappa_{\rho}(v) := \frac{\rho_1(v) + 1}{v} \text{ if } v \neq 0, \kappa_{\rho}(0) = -1$$

Interestingly, the form of  $\rho(\cdot)$  for EL leads to  $\pi_{\rho, i, n}(\theta) = \kappa_{\rho, i, n}(\theta)$  for  $i = 1, \dots, n$ . It is because of this along with the orthogonalization property of the implied probabilities  $\pi_{\rho, i, n}(\theta)$  (shown in Proposition 2 below) that the EL estimator has superior higher-order properties among the GEL class (see Newey and Smith (2004)).

Note that Assumption  $\rho$ (iii) is a technical assumption needed only for the proofs. Now, we are able to establish some important results relative to the GEL implied probabilities.

**Proposition 1** *Let Assumptions 1, 2, and  $\rho$  hold. Then for  $\theta \in \Theta_n$ :*

(A)  $\lambda_{\rho, n}(\theta)$  defined in (5) is such that  $\lambda_{\rho, n}(\theta) = -\hat{\Omega}_n^{-1}(\theta) \bar{g}_n(\theta) + o_P(n^{-1/2})$ ,



(B)  $\pi_{\rho,i,n}(\theta)$  defined in (6) is such that for a given  $i = 1, \dots, n$ ,

$$\pi_{\rho,i,n}(\theta) = \pi_{EEL,i,n}(\theta) + o_P\left(n^{-3/2}\right)$$

where  $\pi_{EEL,i,n}(\theta)$ 's are the implied probabilities from EEL with the closed-form expression

$$\pi_{EEL,i,n}(\theta) = \frac{1}{n} \left[ 1 - (g(W_i; \theta) - \bar{g}_n(\theta))' \hat{\Omega}_n^{-1}(\theta) \bar{g}_n(\theta) \right] = \frac{1}{n} + O_P\left(n^{-3/2}\right).$$

**Remark:** It follows from (B) that the difference between the EEL and GEL implied probabilities is of a smaller order than that between the EEL implied probabilities and the naive empirical probabilities  $\{1/n\}$ . It may be tempting to argue that the use of the GEL implied probabilities to reweight observations results in an equivalence up to one higher order. However, this result, in itself, is not sufficient for such a claim because (B) is not uniform in  $i = 1, \dots, n$ . We provide a formal proof of this claim in Proposition 2.

**Proposition 2** *Let Assumptions 1, 2, and  $\rho$  hold and let  $\theta$  be an arbitrary element of  $\Theta_n$ . Consider  $n$  i.i.d. realizations  $\{Y_{1,n}, \dots, Y_{n,n}\}$  of a  $d \times 1$  random vector  $Y_n$ . Denote  $\bar{Y}_n = \sum_{i=1}^n Y_{i,n}/n$ . Assume that:  $\bar{Y}_n - E[\bar{Y}_n] \xrightarrow{P} 0$ ,  $\frac{1}{n} \sum_{i=1}^n (Y_{i,n} - \bar{Y}_n) [(g(W_i; \theta) - \bar{g}_n(\theta))', Y'_{i,n}] \xrightarrow{P} [\Omega_{Yg}, \Omega_{YY}]$  (finite) and that*

$$\begin{pmatrix} \sqrt{n}(\bar{Y}_n - E[\bar{Y}_n]) \\ \sqrt{n}(\bar{g}_n(\theta) - E[\bar{g}_n(\theta)]) \end{pmatrix} \xrightarrow{d} N\left(0, \begin{bmatrix} \Omega_{YY} & \Omega_{Yg} \\ \Omega'_{Yg} & V \end{bmatrix}\right).$$

Then, as  $n \rightarrow \infty$ , we have

$$(A) \begin{pmatrix} \sqrt{n} \sum_{i=1}^n \pi_{EEL,i,n}(\theta) (Y_i - E[\bar{Y}_n]) \\ \sqrt{n}(\bar{g}_n(\theta) - E[\bar{g}_n(\theta)]) \end{pmatrix} \xrightarrow{d} N\left(0, \begin{bmatrix} \Omega_{YY} - \Omega_{Yg}V^{-1}\Omega'_{Yg} & 0 \\ 0 & V \end{bmatrix}\right),$$

$$(B) \sqrt{n} \sum_{i=1}^n \pi_{\rho,i,n}(\theta) (Y_i - E[\bar{Y}_n]) - \sqrt{n} \sum_{i=1}^n \pi_{EEL,i,n}(\theta) (Y_i - E[\bar{Y}_n]) \xrightarrow{P} 0.$$

**Remarks:** The proofs of Propositions 1 and 2 are given in Appendix. Some of these results were already established in Chaudhuri and Renault (2020). However, the result for the ET estimator is not covered in Chaudhuri and Renault (2020). This is important because ET is the only GEL estimator fully consistent with Shannon's entropy.

Proposition 2 shows that the weighted average involving implied probabilities is asymptotically independent of the average  $\bar{g}_n(\theta)$ . Replacing  $Y_i$  by the first derivative of  $g(W_i; \theta)$  or by  $g(W_i; \theta)g(W_i; \theta)^T$ , one can deduce that the implied probability estimates of the Jacobian and

variance are asymptotically independent of  $\bar{g}_n(\theta)$ . In the case of weak identification, this asymptotic independence of the estimated Jacobian (and estimated variance) with the moment vector leads to better finite-sample properties.

It follows from Proposition 2 that the use of the implied probabilities provides a more precise estimator of  $E[Y]$  since the asymptotic variance is smaller than  $Var(Y)$ . The score test for subsets of parameters that we will discuss now allows for weak identification that makes the use of implied probabilities necessary. Chaudhuri and Zivot (2011) followed Kleibergen (2005) and therefore implicitly used the EEL (Euclidean Empirical Likelihood) implied probabilities. Our paper opens up the possibility of using other implied probabilities for the same test for subsets of parameters, and demonstrates using simulations that other implied probabilities, such as those from EL, can provide significant improvement in its finite-sample performance.

### 3 Score test for subsets of parameters using the implied probabilities

#### 3.1 Score vector and score statistic using the implied probabilities

Following Chaudhuri and Renault (2020), we define the general score vector:

$$l_n(\theta, \pi^G(\theta), \pi^V(\theta)) = \left\{ \sum_{i=1}^n \pi_{i,n}^G(\theta) G'_i(\theta) \right\} \left[ \sum_{i=1}^n \pi_{i,n}^V(\theta) V_{i,n}(\theta) \right]^{-1} \sqrt{n} \bar{g}_n(\theta) \quad (7)$$

where

$$G_i(\theta) = \frac{\partial g(W_i; \theta)}{\partial \theta'}, \quad V_{i,n}(\theta) = g(W_i; \theta) (g(W_i; \theta) - \bar{g}_n(\theta))',$$

and  $\pi_{i,n}^G(\theta)$  and  $\pi_{i,n}^V(\theta)$  may be different, but such that:

$$\pi_{i,n}^G(\theta), \pi_{i,n}^V(\theta) \in \left\{ \hat{\pi}_{i,n}^{(\gamma)}(\theta); \gamma \in \mathbb{R} \right\} \cup \left\{ \frac{1}{n} \right\}. \quad (8)$$

The choice of  $\pi_{i,n}^G(\theta) = \pi_{i,n}^V(\theta) = 1/n$  leads to the standard GMM score statistic (3) as defined in Newey and West (1987). The choice of  $\pi_{i,n}^G(\theta) = \hat{\pi}_{i,n}^{(1)}(\theta)$  (EEL) and  $\pi_{i,n}^V(\theta) = 1/n$  leads to Kleibergen (2005)'s K-statistic. The other choices in (8) cover the various score statistics of Guggenberger and Smith (2005). Importantly, note that  $\pi_{i,n}^G(\theta)$  and  $\pi_{i,n}^V(\theta)$  can be based on different  $\gamma$ s, accommodating for hybrid GEL score statistics in the spirit of Schennach (2007). We refer the interested reader to Chaudhuri and Renault (2020) for further discussion on the score vector.

Pretending that the parameters are all strongly identified, the natural estimator of the asymptotic variance of  $l_n(\theta, \pi^G(\theta), \pi^V(\theta))$  would be

$$\mathcal{I}_n(\theta, \pi^G(\theta), \pi^V(\theta)) = \left\{ \sum_{i=1}^n \pi_{i,n}^G(\theta) G'_i(\theta) \right\} \left[ \sum_{i=1}^n \pi_{i,n}^V(\theta) V_{i,n}(\theta) \right]^{-1} \left\{ \sum_{i=1}^n \pi_{i,n}^G(\theta) G_i(\theta) \right\}. \quad (9)$$

Using (9), the general score statistic based on the general score vector in (7) is given by

$$\mathcal{LM}_n(\theta, \pi^G(\theta), \pi^V(\theta)) = l'_n(\theta, \pi^G(\theta), \pi^V(\theta)) \mathcal{I}_n^{-1}(\theta, \pi^G(\theta), \pi^V(\theta)) l_n(\theta, \pi^G(\theta), \pi^V(\theta)). \quad (10)$$

It is now well-known that if  $\theta_2$  is weakly identified then plugging in a GMM estimator of  $\theta_2$ , that is restricted by  $H_0$  in (2), generally results in badly over-sized test; see Andrews (2017) for a comprehensive discussion. An alternative to such plug-in tests is the projection tests as in, e.g., Dufour and Taamouti (2005, 2007). However, projection tests can be needlessly conservative.

Therefore, we will adopt here the idea of the refined projection score test as in Chaudhuri (2008), Zivot and Chaudhuri (2009), Chaudhuri et al. (2010), Chaudhuri and Zivot (2011). Our presentation can be adapted to the more sophisticated version of the aforementioned tests that were introduced in Andrews (2017), but that is not done here for simplicity and brevity.

To present the refined projection score test for the null hypothesis (2) on  $\theta_1$ , treating  $\theta_2$  as the nuisance parameters, it will be useful to introduce the natural partition of  $l_n(\theta, \pi^G(\theta), \pi^V(\theta))$  and  $\mathcal{I}_n(\theta, \pi^G(\theta), \pi^V(\theta))$  conformable to the partition of  $\theta = (\theta'_1, \theta'_2)'$  as

$$l_n(\theta, \pi^G(\theta), \pi^V(\theta)) = \begin{bmatrix} l_{1,n}(\theta, \pi^G(\theta), \pi^V(\theta)) \\ l_{2,n}(\theta, \pi^G(\theta), \pi^V(\theta)) \end{bmatrix},$$

$$\mathcal{I}_n(\theta, \pi^G(\theta), \pi^V(\theta)) = \begin{bmatrix} \mathcal{I}_{11,n}(\theta, \pi^G(\theta), \pi^V(\theta)) & \mathcal{I}_{12,n}(\theta, \pi^G(\theta), \pi^V(\theta)) \\ \mathcal{I}_{21,n}(\theta, \pi^G(\theta), \pi^V(\theta)) & \mathcal{I}_{22,n}(\theta, \pi^G(\theta), \pi^V(\theta)) \end{bmatrix}, \quad (11)$$

$$l_{1,2,n}(\theta, \pi^G(\theta), \pi^V(\theta)) = l_{1,n}(\cdot, \cdot, \cdot) - \mathcal{I}_{12,n}(\cdot, \cdot, \cdot) \mathcal{I}_{22,n}^{-1}(\cdot, \cdot, \cdot) l_{2,n}(\cdot, \cdot, \cdot),$$

$$\mathcal{I}_{11,2,n}(\theta, \pi^G(\theta), \pi^V(\theta)) = \mathcal{I}_{11,n}(\cdot, \cdot, \cdot) - \mathcal{I}_{12,n}(\cdot, \cdot, \cdot) \mathcal{I}_{22,n}^{-1}(\cdot, \cdot, \cdot) \mathcal{I}_{21,n}(\cdot, \cdot, \cdot)$$

where the right hand side of the last two lines above use  $(\cdot, \cdot, \cdot)$  to denote  $(\theta, \pi^G(\theta), \pi^V(\theta))$  to avoid notational clutter. Using the notation in (11), it is straightforward to decompose the score statistic in (10) as follows:

$$\mathcal{LM}_n(\theta, \pi^G(\theta), \pi^V(\theta)) = \mathcal{LM}_{2,n}(\theta, \pi^G(\theta), \pi^V(\theta)) + \mathcal{LM}_{11,2}(\theta, \pi^G(\theta), \pi^V(\theta)) \quad (12)$$

where, borrowing the maximum-likelihood-terminology from [Cox and Hinkley \(1974\)](#),

$$\begin{aligned}\mathcal{LM}_{2,n}(\theta, \pi^G(\theta), \pi^V(\theta)) &= l'_{2,n}(\theta, \pi^G(\theta), \pi^V(\theta)) \mathcal{I}_{22,n}^{-1}(\theta, \pi^G(\theta), \pi^V(\theta)) l_{2,n}(\theta, \pi^G(\theta), \pi^V(\theta)), \\ \mathcal{LM}_{1.2,n}(\theta, \pi^G(\theta), \pi^V(\theta)) &= l'_{1.2,n}(\theta, \pi^G(\theta), \pi^V(\theta)) \mathcal{I}_{11.2,n}^{-1}(\theta, \pi^G(\theta), \pi^V(\theta)) l_{1.2,n}(\theta, \pi^G(\theta), \pi^V(\theta))\end{aligned}$$

are respectively the score statistic for  $\theta_2$  and the efficient score statistic for  $\theta_1$ . The efficient score statistic  $\mathcal{LM}_{1.2,n}(\theta, \pi^G(\theta), \pi^V(\theta))$  expressed at  $\theta = (\theta_{10}, \theta_2^0)$  can be seen as [Neyman \(1959\)](#)'s  $C(\alpha)$  statistic for testing  $H_0 : \theta_1 = \theta_{10}$ . Interestingly, this test has under standard regularity conditions an asymptotic distribution that is invariant to  $\sqrt{n}$ -local perturbation of  $\theta_2$  from the truth  $\theta_2^0$ ; see, e.g., [Bera and Biliias \(2001\)](#). So the unknown nuisance parameter  $\theta_2^0$  can be replaced by a  $\sqrt{n}$ -consistent estimator without altering the asymptotic distribution of the  $C(\alpha)$  statistic.

Another important fact is that  $\mathcal{LM}_{1.2,n}(\theta, \pi^G(\theta), \pi^V(\theta))$  can be constructed using any choice of implied probabilities (including  $1/n$ ) for the Jacobian or the variance matrix, which will now allow us to explore the improved performance of the refined projection score test idea for the null hypothesis  $H_0 : \theta_1 = \theta_{10}$  in [\(2\)](#) with the use of these implied probabilities.

### 3.2 Refined projection score test using the implied probabilities

To test the null hypothesis  $H_0 : \theta_1 = \theta_{10}$ , we propose to use the refined projection score test, as in [Chaudhuri \(2008\)](#), [Zivot and Chaudhuri \(2009\)](#), [Chaudhuri et al. \(2010\)](#), [Chaudhuri and Zivot \(2011\)](#), but with the accommodation for the various choice of implied probabilities. The test is conducted in two steps:

- **Step 1:** Construct a  $100(1 - \tau)\%$  confidence interval  $\mathcal{C}_{H_0}(\theta_2, 1 - \tau)$  for  $\theta_2$  under the restriction of the null hypothesis  $H_0 : \theta_1 = \theta_{10}$ .  $\mathcal{C}_{H_0}(\theta_2, 1 - \tau)$  is a random subset of the parameter space  $\Theta_2$  of  $\theta_2$  and is defined as follows:

$$\mathcal{C}_{H_0}(\theta_2, 1 - \tau) = \left\{ \theta_2 \in \Theta_2 \mid n\bar{g}_n^T(\theta_{10}, \theta_2) \left[ \hat{\Omega}_n(\theta_{10}, \theta_2) \right]^{-1} \bar{g}_n(\theta_{10}, \theta_2) \leq \chi_{d_g}^2(1 - \tau) \right\}$$

where  $\chi_a^2(b)$  denotes the  $b$ -th quantile of a chi-square distribution with  $a$  degrees of freedom.

- **Step 2:** Reject the null hypothesis  $H_0 : \theta_1 = \theta_{10}$  if either  $\mathcal{C}_{H_0}(\theta_2, 1 - \tau)$  is empty or

$$\inf_{\theta_2 \in \mathcal{C}_{H_0}(\theta_2, 1 - \tau)} \mathcal{LM}_{1.2,n}(\theta_{10}, \theta_2, \pi^G(\theta_{10}, \theta_2), \pi^V(\theta_{10}, \theta_2)) \geq \chi_{d_{\theta_1}}^2(1 - \alpha)$$

where  $d_{\theta_1}$  is the dimension of  $\theta_1$ . When deemed necessary, one should impose  $\pi_{i,n}^G(\theta_{10}, \theta_2) \neq$

$1/n$  following [Kleibergen \(2002, 2005\)](#) to be robust to weak identification of  $\theta$ .

Step 1 corresponds to inverting the S-test of [Stock and Wright \(2000\)](#). In special cases, such as the linear instrumental variables regression with conditionally homoskedastic error,  $\mathcal{C}_{H_0}(\theta_2, 1 - \tau)$  can be obtained analytically using closed-form formula presented in [Dufour and Taamouti \(2005\)](#). Moreover, [Sun \(2018\)](#) provides a STATA command “twostepweakiv” with the “project” option to obtain confidence intervals for  $\theta_1$  based on the version of this refined projection test from [Chaudhuri and Zivot \(2011\)](#).

The difference between the refined projection test and the [Newey and West \(1987\)](#), [Kleibergen \(2002, 2005\)](#) or [Guggenberger and Smith \(2005\)](#) score test is that the former does a projection of  $\mathcal{LM}_{1,2,n}(\cdot)$  from  $\mathcal{C}_{H_0}(\theta_2, 1 - \tau)$  while the latter plugs in an estimator of  $\theta_2$  in  $\mathcal{LM}_n(\cdot)$  that makes  $\mathcal{LM}_{2,n}(\cdot)$  in (12) zero. This difference enables the refined projection test to guard against uncontrolled over-rejection of a true  $H_0$  under weak identification. All these tests are asymptotically equivalent under strong identification thanks to the  $C(\alpha)$  form of  $\mathcal{LM}_{1,2,n}(\cdot)$ .

On the other hand, the refinement provided by the refined projection test over the standard projection test principle is two-fold. First, the projection is done from  $\mathcal{C}_{H_0}(\theta_2, 1 - \tau)$  instead of from  $\Theta_2$ , as is done by the latter. Second, the test statistic and critical values used are  $\mathcal{LM}_{1,2,n}(\cdot)$  and  $\chi_{d_{\theta_1}}^2(\alpha)$  instead of  $\mathcal{LM}_n(\cdot)$  and  $\chi_{d_\theta}^2(\alpha)$ , as is done by the standard projection score test. The restricted projection from  $\mathcal{C}_{H_0}(\theta_2, 1 - \tau)$  instead of from  $\Theta_2$  and the use of the smaller critical values based on the degrees of freedom  $d_{\theta_1}$  instead of  $d_\theta$  of the chi-squared distribution is what makes the refined projection test more powerful than the standard projections tests, see also [Zivot and Chaudhuri \(2009\)](#), [Chaudhuri et al. \(2010\)](#), and [Chaudhuri and Zivot \(2011\)](#).

Without the weak identification problem, the refined projection test is the efficient test in the sense of [Newey and West \(1987\)](#). The standard projection score test is less powerful. In presence of weak identification, both the standard projection score test and the refined projection score test guard against uncontrolled over-rejection of the truth, while the [Newey and West \(1987\)](#), [Kleibergen \(2002, 2005\)](#) or [Guggenberger and Smith \(2005\)](#) score tests do not do so.

The following proposition makes precise the statement about “uncontrolled over-rejection” and “efficient test” made above. For brevity we list the technical assumptions  $\Theta$ , SW and D in the Appendix. These additional assumptions are essential for establishing the asymptotic properties of the refined projection test in [Chaudhuri and Zivot \(2011\)](#) to which we refer the readers for the proof. Then, by appealing to the results in Propositions 1 and 2 that were obtained under Assumptions 1, 2, and  $\rho$ , the results stated in Proposition 3 follow directly.

**Proposition 3** *Let Assumptions 1, 2,  $\rho$ , and the three assumptions  $\Theta$ , SW and D, stated in*

the Appendix, hold. Then we obtain the following results for the refined projection score test using the implied probabilities in (8):

- (i) The asymptotic size of the test cannot exceed  $\alpha + \tau$  for any choice of  $\alpha > 0$  and  $\tau > 0$  with  $\alpha + \tau < 1$  under a restriction in (8) that  $\pi_{i,n}^G(\theta) \in \left\{ \hat{\pi}_{i,n}^{(\gamma)}(\theta); \gamma \in \mathbb{R} \right\}$ .
- (ii) If all elements of  $\theta$  are strongly identified as in [Newey and West \(1987\)](#), and  $\theta_{10} = \theta_1^0 + b/\sqrt{n}$ , then the test with any given  $\tau > 0$ , such that  $\mathcal{C}_{H_0}(\theta_2, 1 - \tau)$  is non empty, is asymptotically equivalent to the infeasible efficient score test that rejects  $H_0 : \theta_1 = \theta_{10}$  if  $\mathcal{LM}_{1.2,n}(\theta_{10}, \theta_2^0, 1/n, 1/n) \geq \chi_{d_{\theta_1}}^2(\alpha)$ .

**Remark:** The tests discussed here involving various implied probabilities have the same first order asymptotic properties as the test in [Chaudhuri and Zivot \(2011\)](#). Indeed, their asymptotic size cannot exceed  $\alpha + \tau$ , and if there is no problem of weak identification then for any choice of  $\tau$  (however small or large) these tests are asymptotically equivalent to the asymptotically efficient infeasible score test with asymptotic size  $\alpha$ . So, with strong identification, the asymptotic size of this test is  $\alpha$  provided the first-step confidence interval is non-empty. The results in [Chaudhuri and Renault \(2020\)](#) suggest that the use of the implied probabilities could lead to better properties in finite samples. This is precisely what we find in the Monte Carlo experiment described below.

## 4 Monte Carlo experiment

The improvement in the finite-sample size properties of tests by the use of implied probabilities is well known. The characterization of the asymptotic size described in Proposition 3(i) of the refined projection test appeals to the Bonferroni inequality applied to the size properties of two full vector score tests. [Guggenberger and Smith \(2005\)](#) and [Chaudhuri and Renault \(2020\)](#) document evidence that the finite-sample size of full vector score tests with various implied probabilities is similar to their nominal level under various scenarios involving different strength of identification. This will be confirmed here in our simulations.

On the other hand, less attention has been paid to the matter of improvement in power; [Chaudhuri and Renault \(2020\)](#) is an exception but only when testing a full vector ( $\theta$ , and not  $\theta_1$ ). However, there is a big difference between the power of a test for the full vector  $\theta$  versus a test for subset of  $\theta$  and the main advantage of the refined projection test concerns its power. Therefore, we will primarily focus on the power properties of the refined projection score test for  $\theta_1$ , compared to that of the plug-in tests. Since the power properties of the plug-in tests

are better understood when parameters  $\theta$  are strongly identified (see [Andrews \(2017\)](#)), we will maintain strong identification of  $\theta$  in this section.

## 4.1 Design

In this section, we examine a model that is not subject to weak identification but is instead affected by large higher-order moments leading to difficult estimation of the variance matrix.

We generate

$$W_i \sim \text{i.i.d Gamma}(\exp(\theta_1^0) = 1, \exp(\theta_2^0) = 2) \text{ for } i = 1, \dots, n$$

where  $\theta_1^0 = \ln(1) = 0$  and  $\theta_2^0 = \ln(2)$ . We exploit the first two moments of the Gamma distribution, i.e.,  $E[W_i] = \exp(\theta_1^0 + \theta_2^0)$  and  $E[W_i^2] = \exp(\theta_1^0 + 2\theta_2^0) + \exp(2\theta_1^0 + 2\theta_2^0)$  to conduct the score tests. Consequently, the moment vector is defined as

$$g(W_i, (\theta_1, \theta_2)) = \begin{bmatrix} W_i - \exp[\theta_1 + \theta_2] \\ W_i^2 - \exp[\theta_1 + 2\theta_2] - \exp[2\theta_1 + 2\theta_2] \end{bmatrix}$$

and it satisfies the moment restrictions in (1) for  $\theta = \theta^0 = (\theta_1^0, \theta_2^0)'$ . The Jacobian does not depend on  $W_i$ , so implied probabilities are not involved in its estimation. The elements of the moment vector  $g(W_i; \theta^0)$  are skewed. Indeed the skewness of the first element is 2, while that of the second element is approximately 6.6. Moreover, the two elements of the moment vector are strongly leptokurtic with fourth moments equal to 144 (kurtosis = 36) and 8687616 (kurtosis = 84.84) respectively. Hence, the estimation of the variance might be problematic and, therefore, appropriate weighting for the estimator of the variance matrix might be crucial.

## 4.2 Results

There is no weak identification issue in this design. Hence, without the fear of over-rejection of the truth, according to the first-order asymptotics, one could plug in the restricted GMM estimator of  $\theta_2$  in the second-step test statistic  $\mathcal{LM}_{1.2,n}(\cdot)$  instead of minimizing the test statistic  $\mathcal{LM}_{1.2,n}(\cdot)$  over values of  $\theta_2$  in the first-step confidence interval. This is similar in spirit to [Newey and West \(1987\)](#)'s score test. Taking advantage of the  $C(\alpha)$  form of  $\mathcal{LM}_{1.2,n}(\cdot)$ 's asymptotic invariance to  $\sqrt{n}$ -local deviation of  $\theta_2$  from  $\theta_2^0$ , we will plug in the computationally convenient restricted GMM estimator of  $\theta_2$  in  $\mathcal{LM}_{1.2,n}(\cdot)$ . We will consider this plug-in version of the score test for three popular choices: (i)  $\pi^G(\cdot) = \pi^V(\cdot) = 1/n$ ; (ii)  $\pi^G(\cdot) = \pi^V(\cdot) = \hat{\pi}^{(1)}(\cdot)$ , i.e., the EEL implied probabilities; and (iii)  $\pi^G(\cdot) = \pi^V(\cdot) = \hat{\pi}^{(-1)}(\cdot)$ , i.e., the EL implied probabilities.

We will similarly consider each of these choices for the refined projection score test with two choices  $\tau = 1\%$  and  $\tau = 5\%$  for the first-step confidence interval. Asymptotic theory says that all tests considered here are asymptotically equivalent and efficient in this case.

To explore the finite-sample properties of the tests, we run 5000 Monte-Carlo trials for the sample sizes  $n = 100$  and 1000. The theoretical size is  $\alpha = 5\%$  for all tests. Table 1 contains the rejection rate of the null  $H_0 : \theta = \theta_{10}$  of all these tests for a grid of deviations from the null, i.e.,  $\theta_{10} - \theta_1^0$ . The columns contain rejection rates for the plug-in score test and our refined test with two values of  $\tau$ ,  $\tau = 1\%$ ,  $5\%$ . The row with  $\theta_{10} - \theta_1^0 = 0$  corresponds to the empirical size of the tests.

nominal level		Plug-in and refined projection score tests with $\tau = 5\%$ and $\tau = 1\%$								
$\alpha = 5\%$		$\pi^G(\cdot) = \pi^V(\cdot) = 1/n$			$\pi^G(\cdot) = \pi^V(\cdot) = \hat{\pi}^{(1)}(\cdot)$ : EEL			$\pi^G(\cdot) = \pi^V(\cdot) = \hat{\pi}^{(-1)}(\cdot)$ : EL		
n	$\theta_{10} - \theta_1^0$	plug-in	$\tau = 5\%$	$\tau = 1\%$	plug-in	$\tau = 5\%$	$\tau = 1\%$	plug-in	$\tau = 5\%$	$\tau = 1\%$
100	-1	99.7	99.3	98.4	97.5	96.3	95.2	99.5	99.4	98.9
100	-0.8	98.6	96.9	93.7	92.5	89.5	87.0	97.6	96.7	95.2
100	-0.6	93.4	87.7	79.2	78.4	72.3	67.3	91.2	88.7	83.5
100	-0.4	76.2	64.0	50.4	50.9	42.9	37.6	72.6	64.9	55.4
100	-0.2	42.1	28.6	18.1	20.3	15.2	12.2	39.1	29.3	21.3
100	0	10.6	6.1	2.9	7.8	5.6	4.6	13.0	7.7	5.4
100	0.2	6.6	6.1	6.0	21.8	18.6	17.9	26.1	20.5	20.1
100	0.4	34.4	34.4	34.4	57.4	50.8	48.5	68.4	63.8	63.6
100	0.6	74.4	74.4	74.4	76.4	66.5	61.6	94.7	93.4	93.4
100	0.8	92.3	92.3	92.3	62.5	49.0	43.1	99.8	99.8	99.8
100	1	96.2	96.2	96.2	37.8	23.2	18.5	100.0	100.0	100.0
1000	-0.3162	99.3	99.3	99.1	99.3	97.6	92.9	97.1	96.1	94.9
1000	-0.253	96.9	96.3	95.7	97.6	93.4	84.7	91.6	88.1	86.1
1000	-0.1897	86.9	84.9	83.1	91.2	81.5	68.4	76.4	71.1	67.1
1000	-0.1265	60.5	57.3	53.8	73.1	56.9	41.4	48.1	42.3	37.5
1000	-0.0632	25.6	23.2	20.6	39.5	24.7	14.8	18.9	14.9	12.4
1000	0	6.2	5.6	4.8	11.9	6.1	3.1	6.8	4.8	4.0
1000	0.0632	11.4	11.4	11.3	15.8	7.4	5.8	21.3	18.8	18.8
1000	0.1265	45.1	45.1	45.1	33.2	16.0	13.0	61.1	58.2	58.2
1000	0.1897	85.6	85.6	85.6	24.8	13.3	11.1	92.0	90.8	90.8
1000	0.253	98.5	98.5	98.5	8.9	5.1	4.3	99.5	99.4	99.4
1000	0.3162	100.0	100.0	100.0	1.5	1.0	0.8	100.0	100.0	100.0

Table 1: Experiment I: Finite-sample rejection rate (in %) of score tests for  $H_{10} : \theta_1 = \theta_{10}$  with nominal level  $\alpha = 5\%$ . Asymptotic size of refined projection test cannot exceed  $\alpha + \tau$ .

First, we analyze the size. We see that the plug-in version of the score test for all three choices of  $\pi^G(\cdot), \pi^V(\cdot)$  over-rejects the true null. Over rejection goes down for the choices  $\pi^G(\cdot) = \pi^V(\cdot) = 1/n$  and  $\pi^G(\cdot) = \pi^V(\cdot) = \hat{\pi}^{(-1)}(\cdot)$  when sample size increases to  $n = 1000$ . However, the refined projection version of the score test for all three choices largely solves this



problem of over-rejection of the truth even when  $n = 100$ . Importantly, we see that the choice of  $\tau = 1\%$  versus  $\tau = 5\%$  for the refined projection does not much affect the finite-sample rejection rate of the truth under this strong identification setup.

Moving to the discussion of power, we see that the refined projection test has a good power in small samples. Now, comparing the choices  $\pi^G(\cdot) = \pi^V(\cdot) = 1/n$ ,  $\pi^G(\cdot) = \pi^V(\cdot) = \hat{\pi}^{(1)}(\cdot)$ ,  $\pi^G(\cdot) = \pi^V(\cdot) = \hat{\pi}^{(-1)}(\cdot)$ , we see that the finite sample-power of the third choice, i.e., EL, is much better than of the other two. The lower power in small samples for the choice  $\pi^G(\cdot) = \pi^V(\cdot) = 1/n$  supports that orthogonalization by the implied probabilities in the variance matrix estimator is important for power. However, do note that the  $\pi^G(\cdot) = \pi^V(\cdot) = \hat{\pi}^{(1)}(\cdot)$  (EEL) delivers the worst power in spite of the orthogonalization by the implied probabilities in the variance matrix estimator. This happens because the EEL implied probabilities can be negative, which rules out the positive (semi-)definiteness of the variance estimator and, in turn, leads to an unduly small  $\mathcal{LM}_{1,2,n}(\cdot)$  under false null hypotheses. The shrinkage of the EEL implied probabilities to make them positive, as suggested in [Antoine et al. \(2007\)](#) and extensively used in [Chaudhuri and Renault \(2020\)](#), can alleviate this problem of poor power to some extent but is not investigated here.

The refined projection test with the EL implied probabilities is the clear winner in terms of size and power. Its superiority is more prominent in the smaller sample where it matters more.

Another Monte Carlo experiment using a linear instrumental variables regression confirms the good size and power of our test<sup>1</sup>.

## 5 Application to the impact of veteran status on earnings

Following [Chaudhuri and Rose \(2009\)](#), we propose to estimate the effect of veteran status on future earnings for the Vietnam war veterans in the United States by running an instrumental variables regression of log annual earnings on the dummy variable veteran status and a variety of control variables related to both earnings and veteran status. One important variable which influences earnings is the years of schooling. However, since schooling is related to some unobservable variable ("ability") that is related to both earnings and veteran status, it is obviously endogenous. So, we wish to estimate a regression of the log earnings on both veteran status and schooling<sup>2</sup>. Given both regressors are endogenous, we need to use instrumental variables.

---

<sup>1</sup>The results are available from the authors upon request.

<sup>2</sup>The causal question in this empirical illustration is a difficult one due to the nature of the relationship between veteran status and schooling. First, veteran status can help increase the years of schooling because of the subsidy provided by the GI Bill. Hence, schooling can be a mediator through which veteran status affects wages. Second,

Angrist (1990, 1991) used the Vietnam Era draft lottery that determined the draft eligibility of individuals, to instrument for an individual’s veteran status in the Vietnam war. A popular choice of instrument for schooling since Card (1995, 1999, 2001) has been the presence of colleges in the neighborhood of where the individual grew up. Following these seminal references, we use four instrumental variables: (i) the lottery number assigned to the individual based on his date of birth, (ii) the lottery ceiling for the year when this individual attained draft age, (iii) a dummy variable indicating the presence of a 4 year accredited public college, and (iv) a dummy variable indicating the presence of a 4 year accredited private college in the neighborhood of the individual’s residence in 1966.

Partialling out the control variables from the system by taking the residuals from a regression of the concerned variables on those controls and the intercept, we will focus on instrumental variables regression model

$$y_i = X_{1i}\theta + X_{2i}\theta_2 + u_i$$

with moment vector

$$g(W_i; \theta) = Z_i(y_i - X_{1i}\theta_1 - X_{2i}\theta_2)$$

where  $y_i, X_{1i}, X_{2i}$  denote the residuals from the regression on the controls and the intercept of the variables log earnings, veteran status, years of schooling respectively, and  $Z$  is the  $4 \times 1$  vector of instruments.

We use the same data<sup>3</sup> as in Chaudhuri and Rose (2009), which was obtained from the National Longitudinal Survey of Young Men. The sample includes 1080 (i.e. 39%) veterans and 1674 non-veterans. In this data set, the instruments are weak for both veteran status and schooling with the first stage  $F$  statistic equal to 8.46 and 2.53 respectively.

Using this data, Chaudhuri and Rose (2009) implemented a variety of plug-in methods, namely the subset-K, subset-KJ and subset-CLR tests, and obtained a significant (at the 5% level) negative effect of veteran status. However, these tests are not reliable in the presence of weak identification as shown by Guggenberger et al. (2012) and Andrews (2017).

The only genuinely weak-identification robust method used in Chaudhuri and Rose (2009) was the so-called subset-Anderson Rubin test proposed by Guggenberger et al. (2012). The subset-AR test led to a 90% confidence interval for the coefficient of the veteran status whose upper bound was approximately .095, signifying that rather large positive effects of veteran

---

the draft avoidance behavior of individuals was often enacted by enrolling in college and thereby increasing years of schooling. That is, the decision to join the military or continuation of schooling were often made simultaneously. A more complete analysis is beyond the scope of this paper.

<sup>3</sup>The dataset is available on [https://saraswata.research.mcgill.ca/MC\\_SC\\_Data.xlsx](https://saraswata.research.mcgill.ca/MC_SC_Data.xlsx)

status —  $100(\exp(.095) - 1) = 9.97\%$  increase in wage — could not be ruled out. The lower bound of the subset-AR confidence interval asymptoted to  $-\infty$ , which was a consequence of weak identification. The inclusion of positive values in the confidence rendered this test inconclusive.

The subset-AR test can be conservative when the effective number of over-identifying restrictions (the number of instruments minus the dimension of  $\theta_2$ , in this case  $4 - 1 = 3$ ) is larger than the number of restrictions in the null (in this case, 1) being tested. Therefore, a priori there is reason to believe that the refined projection test, that is the efficient test under strong identification but also robust to weak identification, might alter the conclusion of the subset-AR test.

Indeed this is what we find with the refined projection test using EL implied probabilities  $\pi^G(\cdot) = \pi^V(\cdot) = \hat{\pi}^{(-1)}(\cdot)$ . This confidence interval also includes implausibly large negative values (consequence of weak identification), however its upper bound is less than zero supporting the hypothesis that the veteran effect is negative.

For a visual illustration, Figure 1 presents two plots against various values of  $\theta_{10}$  of  $H_0 : \theta_1 = \theta_{10}$  — (i) the subset-AR statistic minus the  $\chi_3^2(1 - .1)$ , i.e., the tests statistic minus the 10% critical value for the subset-AR test, and (ii) the second step test statistic for the refined projection test minus the second step critical value, i.e.,  $\inf_{\theta_2 \in \mathcal{C}_{H_0}(\theta_2, 1 - \tau)} \mathcal{LM}_{1.2,n}(\theta_{10}, \theta_2, \pi^G(\theta_{10}, \theta_2), \pi^V(\theta_{10}, \theta_2)) - \chi_1^2(1 - \alpha)$  for the choice:  $\tau = \alpha = .05$ . We take the function plotted for (ii) as  $+\infty$  if the first-step confidence interval is empty (that automatically rejects  $H_0 : \theta_1 = \theta_{10}$  without requiring the second step). The values  $\theta_{10}$  for which these two plots are below the horizontal red dotted line at zero are those that are included in the confidence interval for the respective tests. The vertical black dotted line is the zero effect line. Inclusion of the blue or the green line in the south-east quadrant of the graph means positive effect is not ruled out by the concerned test. We see that while the CI of the subset-AR test includes positive values, that of our refined test includes only negative values which permits to conclude that the veteran effect is negative.

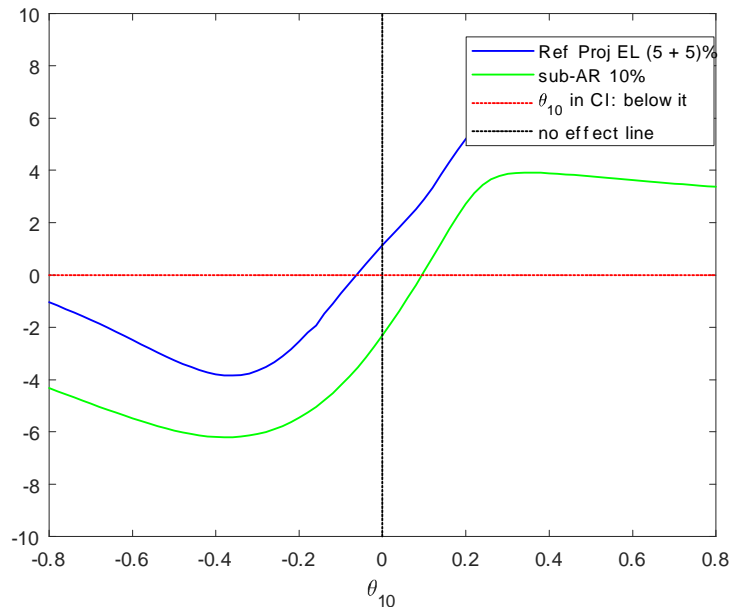


Figure 1: The values of  $\theta_{10}$  below the horizontal line are included in the confidence interval obtained by inverting the refined projection test (blue line) and the subset-AR test (green line).

## 6 Conclusion

In this paper, we propose a two-step approach for testing subvectors of parameters in models characterized by a vector of moment restrictions. The first step is based on an identification robust confidence interval of the parameter while the second relies on a score test. We show the advantages of using implied probabilities obtained from Information Theory criteria to estimate the Jacobian and variance matrix present in our score tests. These tests exploit efficiently the information content of the moment conditions. As a result, these tests have an empirical size close to the theoretical size and their power is good. The resulting confidence intervals are more reliable than those from alternative tests in the presence of skewness and/or weak identification. The theoretical properties of our tests are derived for all the elements of the Cressie-Read family including the Kullback Leibler Information Criterion. Finally, the empirical application brings evidence that veterans have lower earnings than comparable nonveterans.

## 7 Appendix

### 7.1 Assumptions involving weak identification

All the following assumptions are discussed in detail in [Chaudhuri and Zivot \(2011\)](#). Without loss of generality, we group the parameters into weakly and strongly identified parameters. For  $j = w, s$ , let  $\nu_j = \nu_{1j} + \nu_{2j}$ ,  $\theta_j = (\theta'_{1j}, \theta'_{2j})'$  and  $\Theta_j = \Theta_{1j} \times \Theta_{2j}$ . This notation denotes the weakly identified parameters as  $\theta_w$  and the strongly identified parameters as  $\theta_s$ . The true values are, when convenient, regrouped as  $\theta_{0w} = (\theta'_{01w}, \theta'_{02w})'$  and  $\theta_{0s} = (\theta'_{01s}, \theta'_{02s})'$  respectively. When necessary,  $\mathcal{N} \subset \Theta$  and  $\mathcal{N}_r \subset \Theta_r$  are generically used to denote non-shrinking open neighborhoods of  $\theta_0$  and  $\theta_{0r}$  for  $r = 1w, 1s, 2w, 2s, w, s, 1, 2$  respectively. Define  $\tilde{\mathcal{N}} := \mathcal{N}_w \times \mathcal{N}_{1s} \times \Theta_{2s}$ .

**Assumption  $\Theta$ :** [*partition of parameter space*]

For  $l = 1, 2$ , let  $\Theta_l = \Theta_{lw} \times \Theta_{ls}$  and for  $j = w, s$ , let  $\theta_{lj}^0 \in \text{interior}(\Theta_{lj})$  where  $\Theta_{lj} \subset \mathbb{R}^{\nu_{lj}}$  is compact.

**Assumption SW:** [*characterization of strong/weak identification*]

$E[\tilde{g}_n(\theta)] = \tilde{m}_n(\theta)/\sqrt{n} + m(\theta_s)$  where

- (a)  $\tilde{m}_n(\theta) : \Theta \mapsto \mathbb{R}^{d_g}$  is such that  $\tilde{m}_n(\theta) \rightarrow \tilde{m}(\theta)$  uniformly for  $\theta \in \tilde{\mathcal{N}}$  where  $\tilde{m}(\theta)$  is bounded and continuous and  $\tilde{m}(\theta_0) = 0$ . For  $\theta \in \tilde{\mathcal{N}}$ ,  $\tilde{M}_n(\theta) := \partial \tilde{m}_n(\theta) / \partial \theta'$ ,  $\tilde{M}_n(\theta) \rightarrow \tilde{M}(\theta)$  uniformly.  $\tilde{M}(\theta) = [\tilde{M}_{1w}(\theta), \tilde{M}_{1s}(\theta), \tilde{M}_{2w}(\theta), \tilde{M}_{2s}(\theta)]$  where, for  $l = 1, 2$  and  $j = w, s$ , the  $k \times \nu_{lj}$  matrix  $\tilde{M}_{lj}(\theta)$  is bounded and continuous.
- (b)  $m(\theta_s) : \Theta_s \mapsto \mathbb{R}^{d_g}$  is a continuous function and  $m(\theta_s) = 0$  if and only if  $\theta_s = \theta_s^0$ . For  $\theta_s \in \mathcal{N}_{1s} \times \Theta_{2s}$ ,  $M(\theta_s) := \partial m(\theta_s) / \partial \theta'_s$  is bounded and continuous.  $M(\theta_{0s})$  has full column rank. Here,  $M(\theta_s) = [M_1(\theta_s), M_2(\theta_s)]$  where  $M_l(\theta_s) := \partial m(\theta_s) / \partial \theta'_{ls}$  for  $l = 1, 2$ .

**Assumption D:** [*assumptions on the moment vector and its derivative*]

- D1.  $\bar{G}_n(\theta) := \partial \bar{g}_n(\theta) / \partial \theta' = [G_{1wn}(\theta), G_{1sn}(\theta), G_{2wn}(\theta), G_{2sn}(\theta)] = E[\bar{G}_n(\theta)] + o_p(1)$  uniformly for  $\theta \in \tilde{\mathcal{N}}$  where  $E[\bar{G}_n(\theta)] = \partial E[\bar{g}_n(\theta)] / \partial \theta' = \tilde{M}_n(\theta) / \sqrt{n} + [0, M_1(\theta_s), 0, M_2(\theta_s)]$  by imposing interchangeability of the order of differentiation and integration (and from Assumption SW).

D2.  $\sqrt{n} [\bar{g}'_n(\theta_0), \text{vec}'(\bar{G}_{wn}(\theta_0) - E[\bar{G}_{wn}(\theta^0)])] \xrightarrow{d} [\Psi_g^T, \Psi_w^T]$  where <sup>4</sup>

$$\begin{bmatrix} \Psi_g \\ \Psi_w \end{bmatrix} \sim \mathcal{N} \left( 0, \Sigma(\theta_0) = \begin{bmatrix} \Sigma_{gg}(\theta^0) \equiv V(\theta^0) & \Sigma_{gw}(\theta^0) \\ k \times k & k \times k\nu_w \\ \Sigma_{wg}(\theta^0) & \Sigma_{ww}(\theta^0) \\ k\nu_w \times k & k\nu_w \times k\nu_w \end{bmatrix} \right).$$

$\Sigma_{gg}(\theta)$  is bounded, continuous and positive definite. Refining Assumption 2 and with the obvious correspondence of notation between the  $V$ 's and the  $\Sigma$ 's that make precise what the estimators are, we also make the following assumptions.  $\hat{\Sigma}_{gg}(\theta) \xrightarrow{p} \Sigma_{gg}(\theta)$  uniformly for  $\theta \in \tilde{\mathcal{N}}$ .  $\Sigma_{wg}(\theta)$  is bounded and continuous.  $\hat{\Sigma}_{wg}(\theta) := [\hat{\Sigma}'_{1g}(\theta), \dots, \hat{\Sigma}'_{\nu_{1w},g}(\theta), \hat{\Sigma}'_{\nu_1+1,g}(\theta), \dots, \hat{\Sigma}'_{\nu_1+\nu_{2w},g}(\theta)]' \xrightarrow{p} \Sigma_{wg}(\theta)$  uniformly for  $\theta \in \mathcal{N}$ .<sup>5</sup> For  $l = \nu_{1w} + 1, \dots, \nu_1, \nu_1 + \nu_{2w} + 1, \dots, \nu$  the  $d_g \times d_g$  matrices  $\hat{\Sigma}_{lg}(\theta)$  are such that  $\hat{\Sigma}_{lg}(\theta)\hat{\Sigma}_{gg}^{-1}(\theta) = O_p(1)$  uniformly for  $\theta \in \mathcal{N}$ .

## 7.2 Proofs

In the proofs, we use the notation  $g_i(\theta) = g(W_i; \theta)$ .

### Proof of Proposition 1:

(A) A mean-value expansion of the RHS of the (approximate) first-order condition of the maximization problem in (5) gives,

$$\begin{aligned} o_P \left( \frac{1}{\sqrt{n}} \right) &= \frac{1}{n} \sum_{i=1}^n \rho_1(\lambda'_{\rho,n}(\theta) g_i(\theta)) g_i(\theta) \\ &= \frac{1}{n} \sum_{i=1}^n \rho_1(0) g_i(\theta) + \frac{1}{n} \sum_{i=1}^n \rho_2(0) g_i(\theta) g'_i(\theta) \lambda_{\rho,n}(\theta) + R_{\lambda,n}(\theta) \\ &= -\bar{g}_n(\theta) - \hat{\Omega}_n(\theta) \lambda_{\rho,n}(\theta) + R_{\lambda,n}(\theta), \end{aligned} \quad (13)$$

where  $\bar{v}_i$  are the mean-values satisfying  $|\bar{v}_i| \leq |\lambda'_{\rho,n}(\theta) g_i(\theta)|$  for all  $i = 1, \dots, n$ , and the remainder term  $R_{\lambda,n}(\theta) = \frac{1}{n} \sum_{i=1}^n [\rho_2(\bar{v}_i) - \rho_2(0)] g_i(\theta) g'_i(\theta) \lambda_{\rho,n}(\theta)$ . If we could ignore the contribution of  $R_{\lambda,n}(\theta)$  in (13), we would get

$$\lambda_{\rho,n}(\theta) = -\hat{\Omega}_n^{-1}(\theta) \bar{g}_n(\theta) + \hat{\Omega}_n^{-1}(\theta) \times o_P \left( \frac{1}{\sqrt{n}} \right) = o_P \left( \frac{1}{\sqrt{n}} \right).$$

since  $\hat{\Omega}_n(\theta)$  and  $\hat{\Omega}_n^{-1}(\theta)$  are assumed to be  $O_P(1)$  by assumption A2 and  $\bar{g}_n(\theta) = O_P(n^{-1/2})$  by assumption A1 (i) and (iv). For this reason, if we can show that  $\|R_{\lambda,n}(\theta)\| = o_P(n^{-1/2})$ , then

<sup>4</sup>The partition of  $\Psi_w(\theta) = [\Psi'_{1w}(\theta), \Psi'_{2w}(\theta)]'$ ,  $\Sigma_{gw}(\theta) = [\Sigma_{g1}(\theta), \Sigma_{g2}(\theta)] = \Sigma'_{wg}(\theta)$  and  $\Sigma_{ww}(\theta) = (\Sigma_{ll'}(\theta))_{l,l'=1,2}$  is conformable to the partition of  $\theta_w = (\theta'_{1w}, \theta'_{2w})'$ , the partition of the weakly identified elements of  $\theta$  into those from  $\theta_1$  and  $\theta_2$  respectively.

<sup>5</sup>It is worth noting that the  $d_g \times d_g$  matrix  $\hat{\Sigma}_{lg}(\theta^0) \xrightarrow{p} \Sigma_{lg}(\theta^0) = \text{Asym.Cov}(n^{-1/2} \partial \bar{g}_n(\theta^0) / \partial \theta_l, n^{-1/2} \bar{g}_n(\theta^0))$  where  $\theta_l$  is the  $l$ -th element ( $l = 1, \dots, \nu_{1w}, \nu_1 + 1, \dots, \nu_1 + \nu_{2w}$ ) of  $\theta$ .

it will be sufficient to establish result (A). This is what we prove next.

$$\begin{aligned}
\|R_{\lambda,n}(\theta)\| &= \left\| \left( \frac{1}{n} \sum_{i=1}^n [\rho_2(\bar{v}_i) - \rho_2(0)] g_i(\theta) g'_i(\theta) \right) \lambda_{\rho,n}(\theta) \right\| \\
&\leq \left\| \frac{1}{n} \sum_{i=1}^n [\rho_2(\bar{v}_i) - \rho_2(0)] g_i(\theta) g'_i(\theta) \right\| \times \|\lambda_{\rho,n}(\theta)\| \\
&\leq \max_{1 \leq i \leq n} |\rho_2(\bar{v}_i) - \rho_2(0)| \times \left\| \frac{1}{n} \sum_{i=1}^n g_i(\theta) g'_i(\theta) \right\| \times \|\lambda_{\rho,n}(\theta)\| \\
&\leq b \times \max_{1 \leq i \leq n} |\bar{v}_i| \times (\|V(\theta)\| + o_P(1)) \times \|\lambda_{\rho,n}(\theta)\| \\
&\leq b \times \max_{1 \leq i \leq n} |g'_i(\theta) \lambda_{\rho,n}(\theta)| \times b_{\max}(\theta) \times \|\lambda_{\rho,n}(\theta)\| \\
&\leq b \times \max_{1 \leq i \leq n} \|g_i(\theta)\| \times b_{\max}(\theta) \times \|\lambda_{\rho,n}(\theta)\|^2 \\
&\leq b \times b_{\max}(\theta) \times o_P(\sqrt{n}) \times \|\lambda_{\rho,n}(\theta)\|^2 = o_P\left(n^{-1/2}\right), \tag{14}
\end{aligned}$$

by repeated use of Cauchy-Schwartz and triangle inequalities and because  $\max_{1 \leq i \leq n} \|g_i(\theta)\| = o_P(\sqrt{n})$  and  $\|\lambda_{\rho,n}(\theta)\| = O_P(n^{-1/2})$ . Therefore, result (A) follows.

(B) Expanding the numerator and denominator of the RHS of (6) around 0, and using the result obtained in (A), we obtain for any given  $i = 1, \dots, n$

$$\begin{aligned}
\pi_{\rho,i,n}(\theta) &= \frac{\frac{1}{n} [\rho_1(0) + \rho_2(0) \lambda'_{\rho,n}(\theta) g_i(\theta) + \{\rho_2(\bar{v}_i) - \rho_2(0)\} \lambda'_{\rho,n} g_i(\theta)]}{\frac{1}{n} \sum_{j=1}^n [\rho_1(0) + \rho_2(0) \lambda'_{\rho,n} g_j(\theta) + \{\rho_2(\bar{v}_j) - \rho_2(0)\} \lambda'_{\rho,n} g(W_j; \theta)]} \\
&= \frac{\frac{1}{n} [\rho_1(0) - \rho_2(0) g'_i(\theta) \{\hat{\Omega}_n^{-1}(\theta) \bar{g}_n(\theta) + o_P(n^{-1/2})\}] + \{\rho_2(\bar{v}_i) - \rho_2(0)\} \lambda'_{\rho,n}(\theta) g_i(\theta)}{\frac{1}{n} \sum_{j=1}^n [\rho_1(0) - \rho_2(0) g'_j(\theta) \{\hat{\Omega}_n^{-1}(\theta) \bar{g}_n(\theta) + o_P(n^{-1/2})\}] + \{\rho_2(\bar{v}_j) - \rho_2(0)\} \lambda'_{\rho,n}(\theta) g(W_j; \theta)} \\
&= \frac{\frac{1}{n} [1 - (g_i(\theta) - \bar{g}_n(\theta))' \hat{\Omega}_n^{-1}(\theta) \bar{g}_n(\theta)] + R_{NUM,i,n}}{1 - \bar{g}'_n(\theta) \hat{\Omega}_n^{-1}(\theta) \bar{g}_n(\theta) + R_{DEN,n}} \tag{15}
\end{aligned}$$

where the remainder terms in the numerator and the denominator are given by

$$\begin{aligned}
R_{NUM,i,n} &:= \frac{1}{n} \{\rho_2(\bar{v}_i) - \rho_2(0)\} \lambda'_{\rho,n}(\theta) g_i(\theta) - \frac{1}{n} \rho_2(0) g'(W_i; \theta) \times o_P(n^{-1/2}) + \frac{1}{n} \bar{g}'_n(\theta) \hat{\Omega}_n^{-1}(\theta) \bar{g}_n(\theta), \\
R_{DEN,n} &:= \frac{1}{n} \sum_{j=1}^n \left[ \{\rho_2(\bar{v}_j) - \rho_2(0)\} \lambda'_{\rho,n}(\theta) g(W_j; \theta) - \rho_2(0) g'(W_j; \theta) \times o_P(n^{-1/2}) \right].
\end{aligned}$$

It is important to note that  $i$  is given (fixed) in the remainder term  $R_{NUM,i,n}$ . Now following

the same steps as in (A) to deal with the remainder term, we get for a given  $i = 1, \dots, n$

$$\begin{aligned}
|R_{NUM,i,n}| &\leq \frac{1}{n} |\rho_2(\bar{v}_i) - \rho_2(0)| \times \|\lambda_{\rho,n}(\theta)\| \times \|g_i(\theta)\| + \frac{1}{n} \|g_i(\theta)\| \times o_P\left(\frac{1}{\sqrt{n}}\right) + \frac{1}{n} \bar{g}'_n(\theta) \hat{\Omega}_n^{-1}(\theta) \bar{g}_n(\theta) \\
&\leq \frac{1}{n} b \times |\lambda'_{\rho,n}(\theta) g_i(\theta)| \times \|\lambda_{\rho,n}(\theta)\| \times \|g_i(\theta)\| + \|g_i(\theta)\| \times o_P\left(\frac{1}{n^{3/2}}\right) + \frac{1}{n} \|\bar{g}_n(\theta)\|^2 \times b_{min}^{-1}(\theta) \\
&\leq \frac{1}{n} b \times \|\lambda_{\rho,n}(\theta)\|^2 \times \|g_i(\theta)\|^2 + \|g_i(\theta)\| \times o_P\left(\frac{1}{n^{3/2}}\right) + \frac{1}{n} \|\bar{g}_n(\theta)\|^2 \times b_{min}^{-1}(\theta), \\
&= O_P\left(\frac{1}{n^{1+1}}\right) \times O_P(1) + O_P(1) \times o_P\left(\frac{1}{n^{3/2}}\right) + O_P\left(\frac{1}{n^{1+1}}\right) \\
&= o_P\left(\frac{1}{n^{3/2}}\right)
\end{aligned} \tag{16}$$

because  $\|g_i(\theta)\| = O_P(1)$  by A1(iii),  $\|\bar{g}_n(\theta)\| = O_P(n^{-1/2})$  by A1 (i) and (iv), and  $\lambda_{\rho,n}(\theta) = O_P(n^{-1/2})$  by (A). Finally, we want to derive the order of magnitude of  $|R_{DEN,n}|$ . Using a similar technique as before, we obtain that

$$\begin{aligned}
|R_{DEN,n}| &\leq \frac{1}{n} \left| \sum_{j=1}^n \{\rho_2(\bar{v}_j) - \rho_2(0)\} \lambda'_{\rho,n}(\theta) g(W_j; \theta) \right| + \|\bar{g}_n(\theta)\| \times o_P(n^{-1/2}) \\
&\leq \max_{1 \leq j \leq n} |\rho_2(\bar{v}_j) - \rho_2(0)| \times \|\bar{g}_n(\theta)\| \times \|\lambda_{\rho,n}(\theta)\| + \|\bar{g}_n(\theta)\| \times o_P(n^{-1/2}) \\
&\leq b \times \max_{1 \leq j \leq n} |\lambda'_{\rho,n}(\theta) g_j(\theta)| \times \|\bar{g}_n(\theta)\| \times \|\lambda_{\rho,n}(\theta)\| + \|\bar{g}_n(\theta)\| \times o_P(n^{-1/2}) \\
&\leq b \times \max_{1 \leq j \leq n} \|g_j(\theta)\| \times \|\bar{g}_n(\theta)\| \times \|\lambda_{\rho,n}(\theta)\|^2 + \|\bar{g}_n(\theta)\| \times o_P(n^{-1/2}) \\
&= o_P\left(n^{1/2-3/2}\right) + o_P(n^{-1}) = o_P(n^{-1})
\end{aligned}$$

because  $\max_{1 \leq j \leq n} \|g_j(\theta)\| = o_P(\sqrt{n})$  by Assumption 1(ii) while by (A) we have  $\lambda_{\rho,n}(\theta) = O_P(n^{-1/2})$ . Moreover,  $\bar{g}'_n(\theta) \hat{\Omega}_n^{-1}(\theta) \bar{g}_n(\theta)$  in the denominator of (15) is  $O_P(n^{-1})$  because  $\|\bar{g}_n(\theta)\| = O_P(n^{-1/2})$  as before. Therefore, the whole denominator of (15) is  $1 + O_P(n^{-1})$ . Consequently, result (B) follows from (15) and (16).

### Proof of Proposition 2:

(A) This result follows directly from the definition of the  $\pi_{EEL,i,n}(\theta)$  and (vi).

(B) Since our result in Proposition 1(B) is not uniform in  $i$ , we cannot appeal to  $\max_{1 \leq i \leq n} |\pi_{\rho,i,n}(\theta) - \pi_{EEL,i,n}(\theta)|$  after applying Cauchy-Schwartz inequality. Alternatively, we directly work with the expression of the difference  $\{\pi_{\rho,i,n}(\theta) - \pi_{EEL,i,n}(\theta)\} = R_{NUM,i}/(1 + o_P(1))$  obtained in (15). To simplify notations, we denote  $\tilde{Y}_{i,n} := Y_{i,n} - E[\tilde{Y}_n]$ ,  $g_i := g_i(\theta)$ ,  $\bar{g}_n := \bar{g}_n(\theta)$ ,  $\hat{\Omega}_n := \hat{\Omega}_n(\theta)$  and



$\lambda := \lambda_{\rho,n}(\theta)$ . Accordingly, using Proposition 1(A), and assumptions A1 and A2, we obtain

$$\begin{aligned}
& \left\| \sqrt{n} \sum_{i=1}^n \tilde{Y}_{i,n} \{ \pi_{\rho,i,n}(\theta) - \pi_{EEL,i,n}(\theta) \} \right\| \\
\leq & \left\| \frac{1}{n} \sum_{i=1}^n \{ \rho_2(\bar{v}_i) - \rho_2(0) \} \tilde{Y}_{i,n} g'_i \sqrt{n} \lambda \right\| + o_P(1) \times \left\| \frac{1}{n} \sum_{i=1}^n \tilde{Y}_{i,n} g'_i \right\| + \bar{g}'_n \hat{\Omega}_n^{-1} \bar{g}_n \times \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{Y}_{i,n} \right\| \\
\leq & \sqrt{\frac{1}{n} \sum_{i=1}^n \{ \rho_2(\bar{v}_i) - \rho_2(0) \}^2} \times \left\| \frac{1}{n} \sum_{i=1}^n \tilde{Y}_{i,n} g'_i \right\| \times \|\sqrt{n} \lambda\| + o_P(1) \times \left\| \frac{1}{n} \sum_{i=1}^n \tilde{Y}_{i,n} g'_i \right\| + \bar{g}'_n \hat{\Omega}_n^{-1} \bar{g}_n \times \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{Y}_{i,n} \right\| \\
\leq & b \times \sqrt{\frac{1}{n} \sum_{i=1}^n |\lambda' g_i|^2} \times \left\| \frac{1}{n} \sum_{i=1}^n \tilde{Y}_{i,n} g'_i \right\| \times \|\sqrt{n} \lambda\| + o_P(1) \times \left\| \frac{1}{n} \sum_{i=1}^n \tilde{Y}_{i,n} g'_i \right\| + \bar{g}'_n \hat{\Omega}_n^{-1} \bar{g}_n \times \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{Y}_{i,n} \right\| \\
\leq & b \times \sqrt{\frac{1}{n} \sum_{i=1}^n \|g_i\|^2} \times \left\| \frac{1}{n} \sum_{i=1}^n \tilde{Y}_{i,n} g'_i \right\| \times \sqrt{n} \|\lambda\|^2 + o_P(1) \times \left\| \frac{1}{n} \sum_{i=1}^n \tilde{Y}_{i,n} g'_i \right\| + \bar{g}'_n \hat{\Omega}_n^{-1} \bar{g}_n \times \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{Y}_{i,n} \right\| \\
= & O_P(1) \times O_P(1) \times O_P(n^{-1/2}) + o_P(1) \times O_P(1) + O_P(n^{-1}) \times O_P(1) = o_P(1),
\end{aligned}$$

from the standard arguments, for example,  $\|n^{-1} \sum_{i=1}^n \tilde{Y}_{i,n} g'_i\| \leq \|n^{-1} \sum_{i=1}^n \tilde{Y}_{i,n} (g_i - \bar{g}_n)'\| + \|(\bar{Y}_n - E[\bar{Y}_n]) \bar{g}_n\| \leq \|V_{Yg}\| + o_P(1) + O_P(n^{-1/2} \|V_{YY}\|) \times O_P(\|\bar{g}_n\|) = O_P(1)$ .

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Andrews, A., Stock, J., and Sun, L. (2019). Weak instruments in iv regression: Theory and practice. *Annual Review of Economics*, 11:727–753.
- Andrews, D. W. K. (2017). Identification-robust subvector inference. Yale University Working Paper.
- Andrews, I. (2018). Valid two-step identification-robust confidence sets for gmm. *Review of Economics and Statistics*, 100:337–348.
- Angrist, J. (1990). Lifetime earnings and the vietnam era draft lottery: Evidence from social security administrative records. *American Economic Review*, 80:313–336.
- Angrist, J. (1991). The draft lottery and voluntary enlistment in the vietnam era. *Journal of the American Statistical Association*, 86:584–595.
- Antoine, B., Bonnal, H., and Renault, E. (2007). On the efficient use of the informational content

- of estimating equations: Implied probabilities and euclidean empirical likelihood. *Journal of Econometrics*, 138:461–487.
- Bera, A. K. and Biliyas, Y. (2001). Rao’s score, neyman’s  $C(\alpha)$  and silvey’s lm tests: an essay on historical developments and some new results. *Journal of Statistical Planning and Inference*, 97:9–44.
- Caner, M. (2010). Exponential tilting with weak instruments: Estimation and testing. *Oxford Bulletin of Economics and Statistics*, 72:307–326.
- Card, D. (1995). *Using Geographical Variation in College Proximity to Estimate the Return to Schooling. Aspects of Labour Market Behavior: Essays in Honor of John Vanderkamp*. University of Toronto Press.
- Card, D. (1999). *The Causal Effect of Education on Earnings*. Handbook of Labor Economics.
- Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69:1127–1160.
- Chaudhuri, S. (2008). *Projection-Type Score Tests for Subsets of Parameters*. PhD thesis, University of Washington.
- Chaudhuri, S. and Renault, E. (2015). Shrinkage of variance for minimum distance based tests. *Econometric Reviews*, 34:328–351.
- Chaudhuri, S. and Renault, E. (2020). Score tests in gmm: Why use implied probabilities. *Journal of Econometrics*, 219:260–280.
- Chaudhuri, S., Richardson, T., Robins, J. M., and Zivot, E. (2010). Split-sample score tests in linear instrumental variables regression. *Econometric Theory*, 26:1820–1837.
- Chaudhuri, S. and Rose, E. (2009). Estimating the veteran effect with endogenous schooling when instruments are potentially weak. Technical report, University of North Carolina, Chapel Hill and University of Washington.
- Chaudhuri, S. and Zivot, E. (2011). A new method of projection-based inference in gmm with weakly identified nuisance parameters. *Journal of Econometrics*, 164:239–251.
- Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman and Hall.

- Dufour, J. M. (1997). Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica*, 65:1365–1387.
- Dufour, J. M. and Taamouti, M. (2005). Projection-based statistical inference in linear structural models with possibly weak instruments. *Econometrica*, 73:1351–1365.
- Dufour, J. M. and Taamouti, M. (2007). Further results on projection-based inference in iv regressions with weak, collinear or missing instruments. *Journal of Econometrics*, 139:133–153.
- Golan, A. (2006). Information and entropy econometrics — a review and synthesis. *Foundations and Trends in Econometrics*, 2:1–145.
- Golan, A., Judge, G., and Miller, D. (1996). *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. John Wiley & Sons, New York.
- Guggenberger, P., Kleibergen, F., Mavroeidis, S., and Chen, L. (2012). On the asymptotic sizes of subset anderson-rubin and lagrange multiplier tests in linear instrumental variables regression. *Econometrica*, 80:2649–2666.
- Guggenberger, P. and Smith, R. (2005). Generalized empirical likelihood estimators and tests under partial, weak and strong identification. *Econometric Theory*, 21:667–709.
- Imbens, G. W., Spady, R. H., and Johnson, P. (1998). Information theoretic approaches to inference in moment condition models. *Econometrica*, 66:333–357.
- Kitamura, Y. and Stutzer, M. (1997). An information-theoretic alternative to generalized method of moments estimation. *Econometrica*, 65:861–874.
- Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica*, 70:1781–1803.
- Kleibergen, F. (2005). Testing parameters in gmm without assuming that they are identified. *Econometrica*, 73:1103–1123.
- Newey, W. K. and Smith, R. J. (2004). Higher order properties of gmm and generalized empirical likelihood estimators. *Econometrica*, 72:219–255.
- Newey, W. K. and West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 28:777–787.

- Neyman, J. (1959). Optimal asymptotic test of composite statistical hypothesis. In Grenander, U., editor, *Probability and Statistics, the Harald Cramer Volume*, pages 313–334. Almqvist and Wiksell, Uppsala.
- Schennach, S. M. (2007). Point estimation with exponentially tilted empirical likelihood. *The Annals of Statistics*, 35:634–672.
- Smith, R. J. (1997). Alternative semi-parametric likelihood approaches to generalized method of moments estimation. *Economic Journal*, 107:503–519.
- Staiger, D. and Stock, J. (1997). Intrumental variables regression with weak instruments. *Econometrica*, 65:557–586.
- Stock, J. H. and Wright, J. H. (2000). Gmm with weak identification. *Econometrica*, 68:1055–1096.
- Sun, L. (2018). Implementing valid two-step identification-robust confidence sets for linear instrumental-variables model. *The Stata Journal*, 18:803–825.
- Zivot, E. and Chaudhuri, S. (2009). Comment: Weak instrument robust tests in gmm and the new keynesian phillips curve by f. kleibergen and s. mavroeidis. *Journal of Business and Economic Statistics*, 27:328–331.