FUNDAÇÃO GETULIO VARGAS
ESCOLA de PÓS-GRADUAÇÃO em ECONOMIA

Lucas Pimentel Vilela

Hypothesis Testing in Econometric Models

Rio de Janeiro
2017

Lucas Pimentel Vilela

Hypothesis Testing in Econometric Models

Tese submetida à Escola de Pós-Graduação em Economia como requisito parcial para a obtenção do grau de Doutor em Economia

Área de concentração Econometria

Orientador: Marcelo J. Moreira

Rio de Janeiro
2017

Ficha catalográfica elaborada pela Biblioteca Mario Henrique Simonsen/FGV

Vilela, Lucas Pimentel
Hypothesis testing in econometric models / Lucas Pimentel Vilela. - 2017.
76 f.

Tese (doutorado) - Fundação Getulio Vargas, Escola de Pós-Graduação em Economia.
Orientador: Marcelo J. Moreira.
Inclui bibliografia.

1. Testes de hipóteses estatísticas. 2. Variáveis instrumentais (Estatística). 3. Modelos econométricos. I. Moreira, Marcelo J. II. Fundação Getulio Vargas. Escola de Pós- Graduação em Economia. III. Título.

CDD – 519.56

LUCAS PIMENTEL VILELA

"HYPOTHESIS TESTING IN ECONOMETRIC MODELS"

Tese apresentada ao Curso de Doutorado em Economia da Escola de Pós-Graduação em Economia para obtenção do grau de Doutor em Economia.

Data da defesa: 11/12/2015

Aprovada em:

ASSINATURA DOS MEMBROS DA BANCA EXAMINADORA

Marcelo Jovita Moreira
Orientador (a)

Caio Ibsen Rodrigues de Almeida

Humberto Luiz Ataide Moreira

Marcelo Fernandes

Eduardo Fonseca Mendes

Acknowledgements

I would like to express my sincere gratitude and appreciation to my advisor Marcelo J. Moreira, for his guidance and support during my Ph.D program. I also thank my coauthor of the first chapter, Benjamim Mills. I am in indebted to Donald Andrews and James Stock for their help, suggestions and contributions for the first chapter. I wish to thank Jose Diogo, Humberto Moreira, Caio Almeida, Kym Ardison, James Powell, Michael Jansson and Tiago Carvalho for valuable comments on the second and third chapter of this thesis.

This thesis is dedicated to my family: my father Geraldo Vilela Filho, my mother Ana Luiza Pimentel Vilela, my sisters Carolina Pimentel Vilela and Isabela Pimentel Vilela and my girlfriend Thais Buaiz Simao. Without their help and love, this work would not exist. I thank all my friends, who have always been by my side, in particular, my friends from EPGE and my hometown.

Resumo

Esta tese contêm três capítulos. O primeiro capítulo considera testes de hipóteses para o coeficiente de regressão da varíavel endógena em um modelo de variáveis instrumentais. O foco é em testes-t condicionais para hipóteses unilaterais. Trabalhos teóricos e numéricos mostram que os testes-t condicionais centrados nos estimadores de 2SLS e Fuller performam bem mesmo quando os instrumentos são fracamente correlacionados com a variável endógena. Quando a estatística F populacional é menor que dois, o poder é razoavelmente próximo do poder envoltório para testes que são invariantes a transformações que rotacionam os instrumentos (similares ou não similares). Este resultado é surpreendente considerando a baixa performance dos testes-t condicionais para hipóteses bilaterais apresentado em Andrews, Moreira, and Stock (2007). Estes testes possuem baixo poder porque as distribuições das estatísticas-t na hipótese nula são assimétricas quando os instrumentos são fracos. Explorando tal assimetria, nós propomos testes para hipóteses bilaterais baseados em estatísticas-t. Estes testes são aproximadamente não viesados e podem performar tão bem quanto o teste de razão de máxima verossimilhança condicional.

No segundo e no terceiro capítulos, nosso interesse é em testes do tipo maxmin e minimax regret para testes de hipóteses mais gerais. No segundo capítulo, nós apresentamos testes maxmin e minimax regret que satisfazem restrições mais gerais que as restrições de tamanho α e de controle sobre todo o poder na hipótese alternativa. Restrições mais gerais nos possibilitam eliminar testes triviais e obter testes com propriedades desejáveis, como por exemplo não viés, não viés local e similaridade. Na sequência, nós provamos que ambos os testes existem e, sob condições suficientes, eles são testes Bayesianos com *priors* que são solução de um problema de otimização, o problema dual. Na última parte do segundo capítulo, nós consideramos testes de hipóteses que são invariantes à algum grupo de transformações. Sob invariância, o Teorema de Hunt-Stein implica que a busca por testes maxmin e minimax regret pode ser restrita a testes invariantes. Nós provamos que o Teorema de Hunt-Stein continua válido sob as restrições gerais propostas.

No último capítulo, nós desenvolvemos um procedimento numérico para implementar os testes maxmin e minimax regret propostos no segundo capítulo. O espaço paramétrico é discretizado com o objetivo de obter testes de hipóteses com um número finito de pontos. Nós provamos que, ao considerarmos partições mais finas, os testes maxmin e minimax regret que satisfazem um número finito de pontos possuem o mesmo poder na hipótese alternativa que os testes maxmin e minimax regret que satisfazem as restrições gerais. Portanto, nós podemos implementar numericamente os testes que satisfazem um número finito de pontos como aproximação aos testes que satisfazem as restrições gerais.

Palavras-chave: Instrumentos fracos, variáveis instrumentais, testes invariantes, testes maxmin, testes minimax regret, testes não viesados, testes ótimos, testes similares.

Abstract

This thesis contains three chapters. The first chapter considers tests of the parameter of an endogenous variable in an instrumental variables regression model. The focus is on one-sided conditional t-tests. Theoretical and numerical work shows that the conditional 2SLS and Fuller t-tests perform well even when instruments are weakly correlated with the endogenous variable. When the population F-statistic is as small as two, the power is reasonably close to the power envelopes for similar and non-similar tests which are invariant to rotation transformations of the instruments. This finding is surprising considering the poor performance of two-sided conditional t-tests found in Andrews, Moreira, and Stock (2007). These tests have bad power because the conditional null distributions of t-statistics are asymmetric when instruments are weak. Taking this asymmetry into account, we propose two-sided tests based on t-statistics. These novel tests are approximately unbiased and can perform as well as the conditional likelihood ratio (CLR) test.

The second and third chapters are interested in maxmin and minimax regret tests for broader hypothesis testing problems. In the second chapter, we present maxmin and minimax regret tests satisfying more general restrictions than the α-level and the power control over all alternative hypothesis constraints. More general restrictions enable us to eliminate trivial known tests and obtain tests with desirable properties, such as unbiasedness, local unbiasedness and similarity. In sequence, we prove that both tests always exist and under sufficient assumptions, they are Bayes tests with priors that are solutions of an optimization problem, the dual problem. In the last part of the second chapter, we consider testing problems that are invariant to some group of transformations. Under the invariance of the hypothesis testing, the Hunt-Stein Theorem proves that the search for maxmin and minimax regret tests can be restricted to invariant tests. We prove that the Hunt-Stein Theorem still holds under the general constraints proposed.

In the last chapter we develop a numerical method to implement maxmin and minimax regret tests proposed in the second chapter. The parametric space is discretized in order to obtain testing problems with a finite number of restrictions. We prove that, as the discretization turns finer, the maxmin and the minimax regret tests satisfying the finite number of restrictions have the same alternative power of the maxmin and minimax regret tests satisfying the general constraints. Hence, we can numerically implement tests for a finite number of restrictions as an approximation for the tests satisfying the general constraints. The results in the second and third chapters extend and complement the maxmin and minimax regret literature interested in characterizing and implementing both tests.

Keywords: Instrumental variables regression, invariant tests, optimal tests, similar tests, unbiased tests, weak instruments, maxmin tests, minimax regret tests, most stringent tests.

List of Figures

Contents

Chapter 1

Tests Based on t-Statistics for IV Regression with Weak Instruments

1.1 Introduction

Instrumental[1] variables (IVs) are commonly used to make inferences about the coefficient β of an endogenous regressor in a structural equation. When instruments are strongly correlated with the regressor, the tests based on the score (also known as Lagrange Multiplier), likelihood ratio, and t-statistics are asymptotically equivalent. This trinity of tests provides reliable inference as long as the instruments are strong. When identification is weak, however, the three approaches are no longer comparable. Kleibergen (2002) and Moreira (2002) show that a Lagrange Multiplier (LM) statistic has a standard chi-square distribution regardless of the strength of the instruments. Moreira (2003) proposes a conditional likelihood ratio (CLR) test which is shown by Andrews, Moreira, and Stock (2006a) (hereinafter, AMS06a) to be nearly optimal. However, most results in the literature on the performance of tests based on the commonly used t-statistics are negative: Dufour (1997) shows that standard tests based on t-statistics can have size arbitrarily close to one; Andrews, Moreira, and Stock (2006b) (hereinafter, AMS07) find that conditional t-tests are severely biased; and Andrews and Guggenberger (2010) prove that subsampling tests based on the Two-Stage Least Squares (2SLS) t-statistic do not have correct asymptotic size. See Stock, Wright, and Yogo (2002), Dufour (2003), and Andrews and Stock (2007) for surveys on weak IVs.

In this chapter we present conditional one-sided t-tests for testing the null hypothesis $H_0 : \beta = \beta_0$ (or the augmented null $H_0 : \beta \leq \beta_0$) against the alternative $H_1 : \beta > \beta_0$ (the adjustment for $H_1 : \beta < \beta_0$ is straightforward). We consider t-statistics centered around the 2SLS, the limited information maximum likelihood (LIML), the bias-adjusted 2SLS (B2SLS), and the estimator proposed by Fuller (1977) (Fuller's estimator). We also introduce conditional tests based on an one-sided score (LM1) statistic, a likelihood ratio (LR1) statistic for $H_0 : \beta = \beta_0$, and a likelihood ratio statistic (MLR1) for $H_0 : \beta \leq \beta_0$. We develop a theory of optimal tests for one-sided alternatives that parallels the two-sided results of AMS06a. We adopt the same invariance condition as in AMS06a and Chamberlain (2007) under which inference is unchanged if the IVs are transformed by an orthogonal matrix, e.g., by changing the order in which the IVs appear. We develop the Gaussian power envelope for point-optimal invariant similar (POIS) tests. When the null hypothesis is $H_0 : \beta = \beta_0$, the conditional LR1 (CLR1) test is nearly optimal in the sense that its power function is numerically close to the power envelope. For the more relevant null $H_0 : \beta \leq \beta_0$, the CLR1 test does not control size uniformly. The conditional t-tests have correct size and the one based on the 2SLS estimator numerically outperforms the conditional MLR1 (CMLR1) test. The LM1 test is a POIS test and does not have good power overall.

The good performance of the one-sided conditional 2SLS t-test is somewhat surprising considering the poor performance of two-sided conditional t-tests found in AMS07. We show that the poor performance is due to the asymmetric distribution of t-statistics under the null $H_0 : \beta = \beta_0$ when instruments are weak.

[1] The present chapter is an extended version of the paper Mills, Moreira, and Vilela (2014). Some results presented in this chapter are derived in the early work Andrews, Moreira, and Stock (2004). We would like to express our sincere gratitude to Donald Andrews and James Stock for the contributions and suggestions.

We consider two methods to improve power for two-sided tests based on t-statistics. First, we propose novel tests which are by construction approximately unbiased. Second, we modify the t-statistics so that their null distribution is nearly symmetric. Both methods yield some t-tests whose power is close to the CLR test of Moreira (2003). Hence, this chapter restores the triad of tests based on score, likelihood ratio, and t-statistics with reasonably good performance even when instruments are weak for two-sided hypothesis testing. By inverting the conditional t-tests, we can obtain informative confidence regions around different estimators –including the commonly used 2SLS estimator.

The foregoing results are developed under the assumption of normal reduced-form errors with known covariance matrix. The finite-sample theory is extended to non-normal errors with unknown variance at the cost of introducing asymptotic approximations. Under weak instrumental variable (WIV) asymptotics, the exact distributional results extend in large samples to feasible versions of the proposed tests. The finite-sample Gaussian power envelopes are also the asymptotic Gaussian power envelopes with unknown covariance matrix. Under strong-IV asymptotics, we derive consistency even when errors are nonnormal and asymptotic efficiency (AE) when errors are normal[2].

The chapter is organized as follows: section 1.2 introduces the model with one endogenous regressor variable, multiple exogenous regressor variables, and multiple IVs. This section determines sufficient statistics for this model with normal errors and reduced-form covariance matrix. Section 1.3 introduces one-sided invariant similar tests. Section 1.4 focus on the one-sided conditional t-tests. Section 1.5 finds the power envelope for similar and nonsimilar one-sided tests. Section 1.6 adjusts the tests to allow for an estimated error covariance matrix and analyzes their asymptotic properties under weak IVs. Section 1.7 obtains consistency and asymptotic efficiency for one-sided tests. Section 1.8 compares numerically the power of the tests considered in earlier chapters under WIV asympotics. Section 1.9 introduces novel unbiased two-sided tests. Section 1.10 shows that the one-sided conditional t-tests based on 2SLS, LIML and Fuller estimators are asymptotic similar in a uniform sense. Section 1.11 presents confidence intervals for returns to schooling using the data of Angrist and Krueger (1991). An appendix at the end of the thesis contains proofs of the results. The supplement presents: power comparisons for different one-sided and two-sided tests; similar and non-similar power envelopes which are numerically very close (this fact further strengthens our optimality results).

1.2 Model and Sufficient Statistics

In this chapter we study a linear instrumental variable regression model with the objective of making inference about the coefficient of the endogenous variable when the instruments are possibly weak. More specifically, we want to make hypothesis test of β in the following linear model:

$$y_1 = y_2\beta + X\gamma_1 + u, \tag{1.1}$$
$$y_2 = \widetilde{Z}\pi + X\xi_1 + v_2, \tag{1.2}$$

where $y_1, y_2 \in \mathbb{R}^n$, $X \in \mathbb{R}^{n\times p}$, and $\widetilde{Z} \in \mathbb{R}^{n\times k}$ are observed variables; $u, v_2 \in \mathbb{R}^n$ are unobserved errors (possibly correlated); and $\beta \in \mathbb{R}$, $\gamma_1, \xi_1 \in \mathbb{R}^p$, and $\pi \in \mathbb{R}^k$ are unknown parameters. The matrices X and $\widetilde{Z}$ are taken to be fixed (i.e., non-stochastic) and $\overline{Z} := [X : \widetilde{Z}]$ has full column rank $p + k$.

In the first and main part of this chapter we are interested with one-sided hypothesis testing of the coefficient β:[3]

$$H_0 : \beta = \beta_0 (\text{ or } H_0 : \beta \leq \beta_0) \text{ against } H_1 : \beta > \beta_0 \tag{1.3}$$

and in the last part we revise and deal with the two-sided hypothesis testing problem:

$$H_0 : \beta = \beta_0 \text{ against } H_1 : \beta \neq \beta_0 \tag{1.4}$$

It is convenient to transform the IV matrix $\widetilde{Z}$ into a matrix Z which is orthogonal to X: $Z'X = 0$. Since we are interested only in β , we can decompose the IV matrix $\widetilde{Z} = Z + P_X\widetilde{Z} := M_X\widetilde{Z} + P_X\widetilde{Z}$, where

[2] In principle, we could follow Cattaneo, Crump, and Jansson (2012) to obtain efficient one-sided tests when errors are nonnormal, but we do not pursue this line of research here.

[3] The opposite inequality in the null and alternative hypothesis is a straightforward adaptation and we omit the results.

$M_A := I - P_A$ and $P_A := A(A'A)^{-1}A'$ for any full column matrix A, and work with the model:

$$y_1 = y_2\beta + X\gamma_1 + u, \quad (1.5)$$
$$y_2 = Z\pi + X\xi + v_2, \quad (1.6)$$

where $\xi := \xi_1 + (X'X)^{-1}X'\widetilde{Z}\pi$.

Furthermore, the model can be rewritten as a reduced-form in the matrix form:

$$Y = Z\pi a' + X\eta + V, \quad (1.7)$$

where $Y = [y_1 : y_2]$, $V = [v_1 : v_2] := [u + v_2\beta : v_2]$, $a := (\beta, 1)'$, $\eta := [\gamma : \xi]$, and $\gamma := \gamma_1 + \xi\beta$.

The reduced-form errors V are assumed to be independently and identically distributed (i.i.d) across rows. To obtain exact distribution of the tests, we assume that each row has a mean zero bivariate normal distribution with *known* 2×2 nonsingular covariance matrix $\Omega := [\omega_{ij}]_{i,j=1,2}$. As shown below, the normality and the knowledge of Ω assumptions can be relaxed when asymptotic approximations are considered.

The probability model for (1.7) is a member of the curved exponential family, and low dimensional sufficient statistics are available. Lemma 1 of AMS06a shows that $X'Y$ and $Z'Y$ are independent and sufficient for $(\gamma', \xi')'$ and $(\beta, \pi')'$, respectively. Since the assessment of the performance of the tests is by its power we can focus only on tests based on the sufficient statistics, in particular on $Z'Y$. As shown by Moreira (2003), we can apply a one-to-one transformation to $Z'Y$ that yields the $k \times 2$ sufficient statistic $[S : T]$, where:[4]

$$\begin{aligned} S &= (Z'Z)^{-1/2}Z'Yb_0 \cdot (b_0'\Omega b_0)^{-1/2} \text{ and} \\ T &= (Z'Z)^{-1/2}Z'Y\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2}, \end{aligned} \quad (1.8)$$

where $b_0 := (1, -\beta_0)'$ and $a_0 := (\beta_0, 1)'$.

The distribution of the sufficient statistic $[S : T]$ is multivariate normal,

$$vec[S : T] \sim N\left(h_\beta \otimes \mu_\pi, I_{2k}\right), \quad (1.9)$$

with first moment depending on the following quantities:

$$h_\beta := (c_\beta, d_\beta)' \in \mathbb{R}^2 \text{ and } \mu_\pi := (Z'Z)^{1/2}\pi \in \mathbb{R}^k, \quad (1.10)$$

where $c_\beta := (\beta - \beta_0) \cdot (b_0'\Omega b_0)^{-1/2}$ and $d_\beta := a'\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2}$.

1.3 Invariant Similar Tests

Seems natural to suppose that our decision to reject or not the null hypothesis is invariant to changes in the coordinate system of the instrumental variables, i.e the order in which each instrument appears, otherwise there is too much a priori information about the instruments and their relative relevance. The only exception to our knowledge that excludes specific instruments and consequently depends on the order in which the instruments appears is given by Donald and Newey (2001). In consequence we restrict our analysis to tests that are invariant to orthogonal transformations, i.e let ϕ be a $[0, 1]$-valued statistic depending on the sufficient statistics $[S : T]$ and F be a $k \times k$ orthogonal matrix, so the tests considered are such $\phi(FS, FT) = \phi(S, T)$.[5]

By Theorem 6.2.1 of Lehmann and Romano (2005) and Theorem 1 of AMS06a, a test is invariant if and only if it can be written as a function of

$$Q = [S : T]'[S : T] = \begin{bmatrix} S'S & S'T \\ T'S & T'T \end{bmatrix} = \begin{bmatrix} Q_S & Q_{ST} \\ Q_{ST} & Q_T \end{bmatrix}. \quad (1.11)$$

The statistic Q has a Wishart distribution with rank one that depends on

$$\begin{aligned} \xi_\beta(q) &= h_\beta' q h_\beta = c_\beta^2 q_S + 2c_\beta d_\beta q_{ST} + d_\beta^2 q_T, \text{ where} \\ q &= \begin{bmatrix} q_S & q_{ST} \\ q_{ST} & q_T \end{bmatrix} \in \mathbb{R}^{2\times 2}. \end{aligned} \quad (1.12)$$

[4] Henceforth, we use as the matrix square root as the (unique) symmetric square root.

[5] Moreira (2009) shows that the group of transformation on $[S : T]$ is isomorphic to a group of transformations on the original data Y.

Note that $\xi_\beta(q) \geq 0$ because q is positive semi-definite almost surely. The density of Q evaluated at (q_S, q_{ST}, q_T) is given by

$$f_Q(q_S, q_{ST}, q_T; \beta, \lambda) = K_1 \exp(-\lambda(c_\beta^2 + d_\beta^2)/2) \det(q)^{(k-3)/2} \times \exp(-(q_S + q_T)/2)(\lambda\xi_\beta(q))^{-(k-2)/4} I_{(k-2)/2}(\sqrt{\lambda\xi_\beta(q)}), \tag{1.13}$$

where K_1 is a constant, $I_\nu(\cdot)$ denotes the modified Bessel function of the first kind of order ν, and

$$\lambda = \pi' Z' Z \pi \geq 0. \tag{1.14}$$

Examples of invariant test statistics are the Anderson and Rubin (1949), score and likelihood ratio statistics:

$$\begin{aligned} AR &= Q_S/k, \\ LM &= Q_{ST}^2/Q_T, \\ LR &= \frac{1}{2}\left(Q_S - Q_T + \sqrt{(Q_S - Q_T)^2 + 4Q_{ST}^2}\right). \end{aligned} \tag{1.15}$$

When the concentration parameter $\lambda/\left(\omega_{22} \cdot k\right)$ is small, most test statistics are not approximately distributed normal or chi-square. For example, under the weak instrument asymptotics of Staiger and Stock (1997) where $\pi = C/\sqrt{n}$, the LR statistic is not asymptotically pivotal. Its asymptotic distribution is nonstandard and depends on the nuisance and concentration parameter $\lambda/\left(\omega_{22} \cdot k\right)$ under the null. Consequently, the null rejection probability of the standard likelihood ratio test depends on the concentration parameter.

Moreira (2003) proposes similar tests which reject the null hypothesis when the test statistic ψ exceeds a critical value that depends on Q_T:

$$\psi(Q_S, Q_{ST}, Q_T) > \kappa_{\psi,\alpha}(Q_T), \tag{1.16}$$

where $\kappa_{\psi,\alpha}(q_T)$ is the $1-\alpha$ quantile of the distribution of ψ conditional on $Q_T = q_T$ when $\beta = \beta_0$:

$$P_{\beta_0}(\psi(Q_S, Q_{ST}, Q_T)) > \kappa_{\psi,\alpha}(q_T)) = \alpha. \tag{1.17}$$

In practice, the critical value function $\kappa_{\psi,\alpha}(Q_T)$ of the conditional test given in (1.16) is unknown and must be approximated. Given a statistic $\psi(Q_S, Q_{ST}, Q_T)$ write it as a function of Q_S, $\mathcal{S}_2 := Q_{ST}/(||S|| \cdot ||T||)$ and Q_T, by Lemma 3, (f) of AMS06a, $(Q_S, \mathcal{S}_2)$ is independent of Q_T and has a nuisance-parameter free distribution when $\beta = \beta_0$. The null distribution of $(Q_S, \mathcal{S}_2)$ can be approximated by simulating n_{MC} i.i.d random vectors $S_i \sim N(0, I_k)$ for $i = 1, ..., n_{MC}$ where n_{MC} is large. The approximation to $\kappa_{\psi,\alpha}(Q_T)$ is the $1-\alpha$ sample quantile of $\{\psi(S_i'S_i, S_i'e_1 \cdot Q_T^{1/2}, Q_T) : i = 1, ..., n_{MC}\}$ and $e_j \in \mathbb{R}^k$ euclidean vector with one in the j-th coordinate and zero in the others.

We now introduce several new one-sided invariant similar tests for testing $H_0 : \beta = \beta_0$ (or $H_0 : \beta \leq \beta_0$) against $H_1 : \beta > \beta_0$. Each similar test will reject the null hypothesis when the one-sided statistic ψ is larger than the critical value function $\kappa_{\psi,\alpha}$.

The $\underline{k}$-class estimators $\beta\left(\underline{k}\right)$ yield one-sided t-statistics:[6]

$$\begin{aligned} t\left(\underline{k}\right) &= \frac{\beta\left(\underline{k}\right) - \beta_0}{\sigma_u\left(\underline{k}\right)\left[y_2' P_Z y_2 + n\left(1 - \underline{k}\right)\omega_{22}\right]^{-1/2}}, \text{ where} \\ \beta\left(\underline{k}\right) &= \frac{y_2' P_Z y_1 + n\left(1 - \underline{k}\right)\omega_{12}}{y_2' P_Z y_2 + n\left(1 - \underline{k}\right)\omega_{22}} \text{ and } \sigma_u^2\left(\underline{k}\right) = \left(1, -\beta\left(\underline{k}\right)\right)\Omega\left(1, -\beta\left(\underline{k}\right)\right)'. \end{aligned} \tag{1.18}$$

The nonstandard formula for the t-statistics in (1.18) arises here because we take Ω as known (for present purposes only). The commonly used 2SLS estimator, the limited information maximum likelihood for known

[6] To avoid confusion with the number k of exogenous variables, we use $\underline{k}$ to define Theil's class of estimators rather than the more traditional k.

Ω (LIMLK) estimator, the bias-adjusted (B2SLS) estimator proposed by Nagar (1959), and the estimator proposed by Fuller (1977) belong to the $\underline{k}$-class:

$$\begin{array}{ll} \text{2SLS:} & \underline{k} = 1, \\ \text{LIMLK:} & \underline{k} = \underline{k}_{\ LIMLK} = \text{smallest root } \kappa \text{ of } \left|(Y'P_Z Y/n + \Omega) - \kappa\Omega\right| = 0, \\ \text{B2SLS:} & \underline{k} = 1 + (k-2)/n, \text{ and} \\ \text{Fuller:} & \underline{k} = \underline{k}_{\ LIMLK} - 1/n. \end{array} \tag{1.19}$$

The finite-sample properties of the estimators $\beta(\underline{k})$ depend on $\underline{k}$. Consequently, the behavior of the $t(\underline{k})$ statistics can be sensitive to the choice of $\underline{k}$.

We construct two statistics from the likelihood of the model given in (1.7) with Ω known. The first statistic is based on the standard LR statistic (i.e., -2 times the logarithm of the likelihood ratio) for testing $H_0 : \beta = \beta_0$:

$$\begin{aligned} LR1 &= 2\left[\sup_{\beta \geq \beta_0} l_c(Y;\beta,\Omega) - l_c(Y;\beta_0,\Omega)\right] = R(\beta_0) - \inf_{\beta \geq \beta_0} R(\beta), \text{ where} \\ R(\beta) &= \frac{b'Y'P_Z Y b}{b'\Omega b} \text{ with } b = (1,-\beta)', \end{aligned} \tag{1.20}$$

and $l_c(Y;\beta,\Omega)$ is the log-likelihood function for known Ω with all parameters concentrated out except β. In the Appendix, we show that $R(\beta)$ and $LR1$ depend on the observations only through Q defined in (1.11) and

$$LR1 = LR \times 1(\beta(\underline{k}_{\ LIMLK}) \geq \beta_0) + \max\{0, R(\beta_0) - R(\infty)\} \times 1(\beta(\underline{k}_{\ LIMLK}) < \beta_0), \tag{1.21}$$

where $1(\cdot)$ is an indicator function and $R(\infty) = \lim_{\beta\to\infty} R(\beta)$ (hence, $R(\infty)$ equals $R(\beta)$ with b replaced by $(0,-1)'$). We show later that the CLR1 test's power function $P_{\beta,\lambda}(LR1 > \kappa_{LR1,\alpha}(Q_T))$ is not monotonic for $\beta < \beta_0$. Furthermore, the CLR1 test will not have correct size when the null hypothesis is $H_0 : \beta \leq \beta_0$. The second statistic is a standard LR statistic for testing $H_0 : \beta \leq \beta_0$:

$$MLR1 = 2\left[\sup_{\beta} l_c(Y;\beta,\Omega) - \sup_{\beta \leq \beta_0} l_c(Y;\beta,\Omega)\right] = \inf_{\beta \leq \beta_0} R(\beta) - R(\beta(\underline{k}_{\ LIMLK})). \tag{1.22}$$

In the Appendix, we show that

$$\begin{aligned} MLR1 &= [LR - \max\{0, R(\beta_0) - R(\infty)\}] \times 1(\beta(\underline{k}_{\ LIMLK}) \geq \beta_0) \\ &= LR1 - \max\{0, R(\beta_0) - R(\infty)\}. \end{aligned} \tag{1.23}$$

(For $H_1 : \beta < \beta_0$, the inequalities in (1.21) and (1.23) are reversed).

1.4 The Conditional t-Tests

We now elaborate more detailed expressions for the conditional t-tests. It is convenient to write

$$[S:T] = (Z'Z)^{-1/2} Z'Y\Omega^{-1/2} J, \tag{1.24}$$

where J is the orthogonal matrix

$$J = \left[\frac{\Omega^{1/2} b_0}{\sqrt{b_0'\Omega b_0}} : \frac{\Omega^{-1/2} a_0}{\sqrt{a_0'\Omega^{-1} a_0}}\right]. \tag{1.25}$$

From expressions (1.24) and (1.25), we obtain

$$Y'P_Z Y = \Omega^{1/2} J Q J' \Omega^{1/2}, \text{ where } \Omega^{1/2} J = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}. \tag{1.26}$$

We show that the t-statistics are then given by

$$t(\underline{k}) = \frac{\beta(\underline{k}) - \beta_0}{\sigma_u(\underline{k})\left[c_{21}^2 Q_S + 2c_{21}c_{22}Q_{ST} + c_{22}^2 Q_T + n(1-\underline{k})\omega_{22}\right]^{-1/2}}, \text{ where} \tag{1.27}$$
$$\beta(\underline{k}) = \frac{c_{11}c_{21}Q_S + (c_{12}c_{21} + c_{11}c_{22})Q_{ST} + c_{12}c_{22}Q_T + n(1-\underline{k})\omega_{12}}{c_{21}^2 Q_S + 2c_{21}c_{22}Q_{ST} + c_{22}^2 Q_T + n(1-\underline{k})\omega_{22}}.$$

The term $n(1-\underline{k})$ simplifies for each estimator considered in (1.19). Algebraic manipulations show that

$$\begin{array}{ll} \text{2SLS:} & n(1-\underline{k}) = 0, \\ \text{LIMLK:} & n(1-\underline{k}_{\ LIMLK}) = LR - Q_S, \\ \text{B2SLS:} & n(1-\underline{k}) = 2-k, \text{ and} \\ \text{Fuller:} & n(1-\underline{k}) = n(1-\underline{k}_{\ LIMLK}) + 1. \end{array} \tag{1.28}$$

By writing $Q_S = Q_{(k-1)} + LM = Q_{(k-1)} + LM1^2$ and $Q_{ST} = LM1 \cdot Q_T^{1/2}$, we can find the critical value function for each t-statistic as in expression (1.17). For example, the conditional null distribution of the t-statistic based on the 2SLS estimator becomes

$$t(1) = \frac{\beta(\underline{k}) - \beta_0}{\sigma_u(\underline{k})\left[c_{21}^2\left(Q_{(k-1)} + LM1^2\right) + 2c_{21}c_{22}LM1 \cdot q_T^{1/2} + c_{22}^2 q_T\right]^{-1/2}}, \text{ where}$$
$$\beta(1) = \frac{c_{11}c_{21}\left(Q_{(k-1)} + LM1^2\right) + (c_{12}c_{21} + c_{11}c_{22})LM1 q_T^{1/2} + c_{12}c_{22}q_T}{c_{21}^2\left(Q_{(k-1)} + LM1^2\right) + 2c_{21}c_{22}LM1 \cdot q_T^{1/2} + c_{22}^2 q_T}, \tag{1.29}$$

$Q_{(k-1)}$ has a chi-square distribution with $k-1$ degrees of freedom, and $LM1$ has a standard normal distribution.

1.5 Power Envelopes

In this section we address the question of optimal invariant similar and nonsimilar tests when the IV's may be weak. To evaluate the performance of the novel one-sided conditional tests, we derive the power envelopes for similar and nonsimilar tests. The use of sufficiency and invariance reduces the dimension of the parameters from $1+k+2p$ for $\theta = (\beta, \pi', \xi', \gamma')'$ to just 2 for $(\beta, \lambda)'$. The dimension reduction allows the power envelope to meaningfully assess the performance of our one-sided tests. The envelope we derive here consists of upper bound for power and lower bound for either $H_0 : \beta = \beta_0$ or $H_0 : \beta \leq \beta_0$.

1.5.1 Similar Power Envelope

The following theorem is the main result of this section:

Theorem 1 *Define the statistic*

$$LR_{\beta^*\lambda^*}(Q_1, Q_T) = \frac{f_{Q_1,Q_T}(q_1, q_T; \beta^*, \lambda^*)}{f_{Q_T}(q_T; \beta^*, \lambda^*) f_{Q_1|Q_T}(q_1|q_T; \beta_0)} = \frac{\varphi_1(q_1, q_T; \beta^*, \lambda^*)}{\varphi_2(q_T; \beta^*, \lambda^*)}, \tag{1.30}$$

where

$$\varphi_1(q_1, q_T; \beta, \lambda) = \exp(-\lambda c_\beta^2/2)(\lambda \xi_\beta(q))^{-(k-2)/4} I_{(k-2)/2}\left(\sqrt{\lambda \xi_\beta(q)}\right) \text{ and}$$
$$\varphi_2(q_T; \beta, \lambda) = \left(\lambda d_\beta^2 q_T\right)^{-(k-2)/4} I_{(k-2)/2}\left(\sqrt{\lambda d_\beta^2 q_T}\right). \tag{1.31}$$

Let $\kappa_{\beta^*\lambda^*,\alpha}(Q_T)$ be a shorthand for $\kappa_{LR_{\beta^*\lambda^*},\alpha}(Q_T)$. Then the following hold:
(a) *For (β^*, λ^*) with $\beta^* > \beta_0$, the test that rejects $H_0 : \beta = \beta_0$ when $LR_{\beta^*\lambda^*}(Q_1, Q_T) > \kappa_{\beta^*\lambda^*,\alpha}(Q_T)$ maximizes power over all level α invariant similar tests.*
(b) *For (β^*, λ^*) with $\beta^* < \beta_0$, the test that rejects $H_0 : \beta = \beta_0$ when $LR_{\beta^*\lambda^*}(Q_1, Q_T) < \kappa_{\beta^*\lambda^*,1-\alpha}(Q_T)$ minimizes the null rejection probability over all level α invariant similar tests.*

Comments: 1. We denote the test that rejects the null when $LR_{\beta^*\lambda^*}(Q_1, Q_T) > \kappa_{\beta^*\lambda^*,\alpha}(Q_T)$ as a point-optimal invariant similar (POIS) test. We determine the power upper bound by considering the POIS tests for arbitrary values (β^*, λ^*) when $\beta^* > \beta_0$. The power upper bound is for similar tests for $H_0 : \beta = \beta_0$. We do not impose the additional constraint that tests must have correct size, and so the upper bound could be conservative for $H_0 : \beta \leq \beta_0$. We shall see later that even for small values of λ, some tests for $H_0 : \beta \leq \beta_0$ do reach the upper bound.

2. The test which rejects the null when $LR_{\beta^*\lambda^*}(Q_1, Q_T) < \kappa_{\beta^*\lambda^*,1-\alpha}(Q_T)$ is called POIS0 test. We determine the null lower bound by finding the power of POIS0 tests for arbitrary values (β^*, λ^*) when $\beta^* < \beta_0$.

3. The power envelope is the union of the power upper bound and null lower bound. Both bounds are relevant because we would like to compare the probability of making the type I and type II errors for different tests.

4. The denominator $\varphi_2(q_T; \beta^*, \lambda^*)$ does not depend on q_1 and can be absorbed into the conditional critical value. Thus, the test based on $LR_{\beta^*\lambda^*}(Q_1, Q_T)$ is equivalent to a test based on the numerator of $\varphi_1(q_1, q_T; \beta^*, \lambda^*)$. For reasons of numerical stability, however, we recommend constructing critical values using $\ln(LR_{\beta^*\lambda^*}(Q_1, Q_T))$.

We now show that such tests do not depend on λ^*, so that the POIS and POIS0 tests are of a relatively simple form. Using a series expansion of $I_{(k-2)/2}(x)$, we can write

$$\varphi_1(q_1, q_T; \beta, \lambda) = 2^{-(k-2)/2} \exp(-\lambda c_\beta^2/2) \sum_{j=0}^{\infty} \frac{(\lambda \xi_\beta(q_1, q_T)/4)^j}{j!\Gamma((k-2)/2 + j + 1)} \tag{1.32}$$

The term $\varphi_2(q_T; \beta, \lambda)$ can be written analogously.

The function $\varphi_1(q_1, q_T; \beta, \lambda)$ is increasing in $\xi_\beta(q_1, q_T) \geq 0$. As a result, for a fixed value of β, say $\beta^* > \beta_0$, the optimal test for fixed alternative β^* rejects $H_0 : \beta = \beta_0$ when

$$\xi_{\beta^*}(Q_1, Q_T) > \kappa_{\beta^*,\alpha}(Q_T), \tag{1.33}$$

where $\kappa_{\beta^*,\alpha}(Q_T)$ is a shorthand for $\kappa_{\xi_{\beta^*},\alpha}(Q_T)$ as defined in (1.17). This POIS test is *one-sided* because it directs power at a single point β^* that is greater than the null value β_0. An analogous argument shows that the POIS0 test that minimizes rejection probabilities for fixed $\beta^* < \beta_0$ rejects H_0 when

$$\xi_{\beta^*}(Q_1, Q_T) < \kappa_{\beta^*,1-\alpha}(Q_T). \tag{1.34}$$

Corollary 2 *For $\beta^* > \beta_0$, the POIS test based on $\xi_{\beta^*}(Q_1, Q_T)$ is the uniformly most powerful test among invariant similar tests against the alternative distributions indexed by $\{(\beta^*, \lambda) : \lambda > 0\}$. For $\beta^* < \beta_0$, the POIS0 test based on $\xi_{\beta^*}(Q_1, Q_T)$ uniformly minimizes the null rejection probability among invariant similar tests against the alternative distributions indexed by $\{(\beta^*, \lambda) : \lambda > 0\}$.*

Comments: 1. The form of the POIS test depends on the alternative β^*. Hence, there does not exist a Uniformly Most Powerful Invariant (UMPI) test. Although the form of the POIS and POIS0 tests does not depend on λ^*, their power depends on the true value of λ. Hence, the power envelope depends on both parameters β and λ.

2. A test based on $\xi_{\beta^*}(Q_1, Q_T)$ is equivalent to a test that rejects when

$$\begin{aligned} POIS1_\delta &= \frac{Q_S + \delta \mathcal{S}_2 \sqrt{Q_S} - k}{\sqrt{2k + \delta^2}} > \kappa_{\delta,\alpha}(Q_T), \text{ where} \\ \delta &= (2d_{\beta^*}/c_{\beta^*})\sqrt{Q_T},), \end{aligned} \tag{1.35}$$

and $\kappa_{\delta,\alpha}(Q_T)$ is a shorthand for $\kappa_{POIS1_\delta,\alpha}(Q_T)$ defined in (1.17). This formulation of the test is convenient because Q_S, $\mathcal{S}_2$, and Q_T are independent under $\beta = \beta_0$, which simplifies the calculation of critical values.

3. Provided $\omega_{12} - \omega_{22}\beta_0 \neq 0$, the quantity d_{β^*} is a linear function of β^* and equals zero if and only if $\beta^* = \beta_{AR}$, where

$$\beta_{AR} = \frac{\omega_{11} - \omega_{12}\beta_0}{\omega_{12} - \omega_{22}\beta_0}. \tag{1.36}$$

In this case, $\delta = 0$ and $POIS1_\delta$ reduces to $Q_S/\sqrt{2k}$, which is the AR statistic rescaled. Hence, the AR test, usually conceived as a two-sided test, is one-sided POIS against the alternative $\beta = \beta_{AR}$.This finding is in agreement with Chernozhukov, Hansen, and Jansson (2009) who use completeness of Q_T to show that the weighted average power likelihood ratio (WAP-LR) tests of Andrews, Moreira, and Stock (2004) (hereinafter, AMS04) are admissible; see Moreira and Moreira (2013) on admissible WAP-LR similar tests without completeness.

4. The Locally Most Powerful Invariant (LMPI) test is the POIS test for β^* local to β_0 with $\beta^* > \beta_0$. This test is equivalent to the one-sided LM test that rejects H_0 if

$$LM1 = Q_{ST}/Q_T^{1/2} > z_\alpha, \tag{1.37}$$

where z_α is the $1-\alpha$ quantile of the standard normal distribution. Analogously, if β^* is local to β_0 with $\beta^* < \beta_0$, then the LMPI test rejects H_0 if $-Q_{ST}/Q_T^{1/2} > z_\alpha$.

5. The sign of δ in (1.35) can change as β^* changes even for β^* values on the same side of the null hypothesis because d_{β^*} is a linear function of β^*. As a result, the form of the $POIS1_\delta$ statistic (and the power envelope) changes dramatically as β^* varies. The constant δ determines the weight put on the statistic $\mathcal{S}_2$. The optimal value of δ for small values of $\beta > \beta_0$ has the wrong sign for large values of β and vice versa. This fact has adverse consequences for the overall one-sided power properties of POIS tests.

6. The optimal one-sided test for β^* arbitrarily large rejects H_0 if

$$Q_S + 2(\det(\Omega))^{-1/2}(\beta_0\omega_{22} - \omega_{12})Q_{ST} > \kappa_{\infty,\alpha}(Q_T) \tag{1.38}$$

for $\kappa_{\infty,\alpha}(\cdot)$ as defined in (1.17). Remarkably, the same test is the optimal one-sided test for β^* negative and arbitrarily large in absolute value for any λ^*. Consequently, the optimal two-sided test for $|\beta^* - \beta_0|$ arbitrarily large is the test in (1.38).

Corollary 2 shows that the POIS test for an alternative (β^*, λ^*) depends only on β^*. Because the true parameter β is unknown, we could construct an empirical version of the standardized optimal statistic:

$$\widetilde{\xi}_{\widehat{\beta}} = x'_{\widehat{\beta}} Q x_{\widehat{\beta}}, \tag{1.39}$$

where $x_{\widehat{\beta}} = (c_{\widehat{\beta}}/||h_{\widehat{\beta}}||, d_{\widehat{\beta}}/||h_{\widehat{\beta}}||)'$ and $\widehat{\beta}$ is the maximum likelihood estimator of β under $H_1 : \beta > \beta_0$. The next theorem shows that the empirical POIS test is equivalent to those based on $CLR1$ test.

Theorem 3 *The statistics $\widetilde{\xi}_{\widehat{\beta}}$ and $LR1$ are equivalent up to strictly increasing transformations (possibly depending on Q_T). In particular,*

$$P_{\beta,\lambda}(\widetilde{\xi}_{\widehat{\beta}} > \kappa_{\widetilde{\xi}_{\widehat{\beta}},\alpha}(Q_T)) = P_{\beta,\lambda}\left(LR1 > \kappa_{LR1,\alpha}(Q_T)\right). \tag{1.40}$$

Comment: This theorem and Comment 5 of Corollary 2 indicate that the CLR1 test does not have correct size when the null hypothesis is $H_0 : \beta \leq \beta_0$ instead of $H_0 : \beta = \beta_0$. See section 1.8 below for numerical simulations on size and power of the CLR1 test.

1.5.2 Non-Similar Power Envelope

Non-similar tests have null rejection probability below the significance level for some values of the nuisance parameter λ. Due to the continuity of the power function, for such values of λ, the power of a non-similar test is less than the power of a similar test for alternatives close enough to the null hypothesis. However, for other values of λ, or for more distant alternatives, non-similar tests can have greater power than similar tests. For this reason, we also consider optimal invariant non-similar tests of the hypothesis $H_0 : \beta = \beta_0$ against point alternatives.

Our construction of point-optimal invariant (POI) non-similar tests follows Section 3.8 of Lehmann and Romano (2005). Consider the composite null hypothesis

$$H_0 : (\beta, \lambda) \in \{(\beta_0, \lambda) : 0 \leq \lambda < \infty\}, \tag{1.41}$$

and the point alternative

$$H_1 : (\beta, \lambda) = (\beta^*, \lambda^*). \tag{1.42}$$

Let Λ be a probability measure over $\{\lambda : 0 \leq \lambda < \infty\}$ and h_Λ be the weighted pdf,

$$h_\Lambda(q) = \int f_{Q_1,Q_T}(q_1, q_T; \beta_0, \lambda) d\Lambda(\lambda), \tag{1.43}$$

where $f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda)$ is given in (1.13). The effect of weighting by Λ under the null is to turn the composite null into a point null, so that the most powerful test can be obtained using the Neyman-Pearson Lemma. Specifically, let ϕ_Λ be the most powerful test of h_Λ against $f_Q(q; \beta^*, \lambda^*)$, so that ϕ_Λ rejects the null when

$$NP_\Lambda(q) = \frac{f_Q(q; \beta^*, \lambda^*)}{h_\Lambda(q)} > d_{\Lambda,\alpha}, \tag{1.44}$$

where $d_{\Lambda,\alpha}$ is the critical value of the test, chosen so that $NP_\Lambda(q)$ rejects the null with probability α under the distribution h_Λ.

If the test ϕ_Λ has size α for the null hypothesis H_0 in (1.41), i.e.,

$$\sup_{0 \leq \lambda < \infty} P_{\beta_0,\lambda}(NP_\Lambda(Q) > d_{\Lambda,\alpha}) = \alpha, \tag{1.45}$$

then the test ϕ_Λ is most powerful for testing H_0 against H_1, and the distribution Λ is least favorable; cf. Thm. 3.8.1 and Cor. 3.8.1 of Lehmann and Romano (2005).

Given a distribution Λ, condition (1.45) is easily checked numerically. What proves more computationally difficult is finding the distribution that satisfies (1.45). In the numerical work we consider distributions Λ that put point mass on some point λ_0. In this case, we have

$$NP_\Lambda = \frac{f_Q(q; \beta^*, \lambda^*)}{f_Q(q; \beta_0, \lambda_0)} \tag{1.46}$$

Let $\mathcal{R}(\beta_0, \lambda_0, \beta^*, \lambda^* | \beta, \lambda)$ be the rejection probability of the test based on the statistic in (1.46) when the true values are β and λ. The numerical problem is to find the value of λ_0 such that the test has size α. Denote this value of λ_0 by λ_0^{LF}; then λ_0^{LF} solves

$$\begin{aligned} \mathcal{R}(\beta_0, \lambda_0^{LF}, \beta^*, \lambda^* | \beta_0, \lambda_0^{LF}) &= \alpha \text{ and} \\ \sup_{0 \leq \lambda < \infty} \mathcal{R}(\beta_0, \lambda_0^{LF}, \beta^*, \lambda^* | \beta_0, \lambda) &\leq \alpha. \end{aligned} \tag{1.47}$$

If there is a $\lambda_0^{LF}(\beta_0, \beta^*, \lambda^*)$ that satisfies (1.47), then the test based on $NP_{\lambda_0^{LF}}$ is the POI non-similar test.

The power upper bound for invariant non-similar tests is $\mathcal{R}(\beta_0, \lambda_0^{LF}(\beta_0, \beta^*, \lambda^*), \beta^*, \lambda^* | \beta^*, \lambda^*)$ (an analogous argument yields a null lower bound). We find numerically that the power envelopes for similar and non-similar tests are essentially the same, up to numerical accuracy. The reason for this is twofold. On one hand, the conditional critical values for the POIS tests depend on q_T only weakly in the range of q_T that is most likely to occur under the alternative. Thus, the POIS tests are very nearly unconditional. On the other hand, the POI non-similar tests have null rejection rates that are very nearly equal to α for all values of λ; thus, the POI non-similar tests are very nearly similar. Because POI similar tests are nearly unconditional and the POI non-similar tests are nearly similar, the two types of tests have nearly the same rejection regions. An analogous result is described by Andrews, Moreira, and Stock (2008) for two-sided testing.

1.6 Weak IV Asymptotics

Here, we consider the same model and hypotheses as in section 1.2, but with non-normal reduced-form errors with unknown covariance matrix. We show that the finite-sample distribution of the tests and statistics holds asymptotically under the same high-level assumptions as in Staiger and Stock (1997). To model weak IV asymptotics and fixed alternatives (WIV-FA), we let π be local to zero and the alternative β be fixed, not local to the null value β_0.

Assumption WIV-FA. (a) $\pi = C/n^{1/2}$ for some non-stochastic k-vector C.

(b) β is a fixed constant for all $n \geq 1$.
(c) k is a fixed positive integer that does not depend on n.

We now specify the asymptotic behavior of the instruments, exogenous regressors, and reduced-form errors.

Assumption 1. $n^{-1}\overline{Z}'\overline{Z} \to_p D$ for some positive definite $(k+p)\times(k+p)$ matrix D.

Assumption 2. $n^{-1}V'V \to_p \Omega$ for some positive definite 2×2 matrix Ω.

Assumption 3. $n^{-1/2}vec(\overline{Z}'V) \to_d N(0,\Phi)$ for some pd $2(k+p)\times 2(k+p)$ matrix Φ, where $vec(\cdot)$ denotes the column by column vec operator.

Assumption 4. $\Phi = \Omega \otimes D$.

The quantities C, D, and Ω are assumed to be unknown. AMS04 show that Assumptions 1-3 hold under general conditions. Assumption 4 holds under Assumptions 1-3 and homoskedasticity of the errors V_i, i.e., $E(V_iV_i'|\overline{Z}_i) = EV_iV_i' = \Omega$ $a.s.$

We now introduce tests that are suitable for (possibly) non-normal, homoskedastic, uncorrelated errors with unknown covariance matrix. See AMS04 for tests and results for cases in which the errors are not homoskedastic or are correlated. For clarity of the asymptotics results, we write S, T, Q, etc. of section 1.2, as S_n, T_n, Q_n, etc.

1.6.1 Tests for Unknown Ω and Possibly Non-normal Errors

For feasible tests when we do not know the reduced-form error covariance matrix Ω, we need to estimate consistently Ω, and this is achieved by the estimator[7]

$$\widehat{\Omega}_n = n^{-1}\widehat{V}'\widehat{V}, \text{ where } \widehat{V} = M_{Z,X}Y = Y - P_ZY - P_XY, \tag{1.48}$$

see Lemma 1 of Andrews, Moreira, and Stock (2006b) (hereinafter, AMS06b).

Given that, we can replace Ω by $\widehat{\Omega}_n$ and obtain modified versions of the statistics S_n, T_n, $Q_{S,n}$, $Q_{ST,n}$ and $Q_{T,n}$:

$$\begin{aligned}
\widehat{S}_n &= (Z'Z)^{-1/2}Z'Yb_0\cdot(b_0'\widehat{\Omega}_nb_0)^{-1/2},\\
\widehat{T}_n &= (Z'Z)^{-1/2}Z'Y\widehat{\Omega}_n^{-1}a_0\cdot(a_0'\widehat{\Omega}_n^{-1}a_0)^{-1/2},\\
\widehat{Q}_{S,n} &= \widehat{S}_n'\widehat{S}_n,\ \widehat{Q}_{ST,n} = \widehat{S}_n'\widehat{T}_n \text{ and } \widehat{Q}_{T,n} = \widehat{T}_n'\widehat{T}_n.
\end{aligned} \tag{1.49}$$

The feasible one-sided t-statistics for Ω unknown are:

$$\begin{aligned}
\widehat{t}\left(\widehat{\underline{k}}\right)_n &= \frac{\widehat{\beta}(\widehat{\underline{k}}) - \beta_0}{\widehat{\sigma}_u(\widehat{\underline{k}})[y_2'P_Zy_2 + n(1-\widehat{\underline{k}})\widehat{w}_{22}]^{-1/2}}, \text{ where}\\
\widehat{\beta}(\widehat{\underline{k}}) &= \frac{y_2'P_Zy_1 + n(1-\widehat{\underline{k}})\widehat{w}_{12}}{y_2'P_Zy_2 + n(1-\widehat{\underline{k}})\widehat{w}_{22}}, \text{ and } \widehat{\sigma}_u^2\left(\widehat{\underline{k}}\right) = \left(1, -\widehat{\beta}(\widehat{\underline{k}})\right)\widehat{\Omega}_n\left(1, -\widehat{\beta}(\widehat{\underline{k}})\right)'.
\end{aligned}$$

The values of $\widehat{\underline{k}}$ are obtained from (1.28) after estimating Ω :

$$\begin{array}{ll}
\text{2SLS:} & \widehat{\underline{k}} = 1\\
\text{LIML:} & \widehat{\underline{k}} = \widehat{\underline{k}}_{\ LIML} = \text{smallest root of } |(Y'P_ZY/n - \widehat{\Omega}_n) - \kappa\widehat{\Omega}_n| = 0\\
\text{B2SLS:} & \widehat{\underline{k}} = 1 + (k-2)/n\\
\text{Fuller:} & \widehat{\underline{k}} = \widehat{\underline{k}}_{\ LIML} - 1/n.
\end{array} \tag{1.50}$$

The feasible $\widehat{t}\left(\widehat{\underline{k}}\right)_n$ statistics depend on $\widehat{Q}_{1,n}$, $\widehat{Q}_{T,n}$ and $\widehat{\Omega}_n$ in the same way the $t(\underline{k})$ statistics depend on Q_1, Q_T and Ω, as described in previous sections. For all remaining test statistics, we just need to replace Q_S, Q_{ST} and Q_T by their analogues in which Ω is estimated by $\widehat{\Omega}_n$. For example, the $LR1$ test statistic for

[7] This definition of $\widehat{\Omega}_n$ is suitable if Z or X contains a vector of ones, as is usually the case. If not, then $\widehat{\Omega}_n$ is defined with the sample mean of $\widehat{V}$ subtracted from it.

unknown Ω is defined as in (1.21), but with Q_S, Q_{ST} and Q_T replaced by $\widehat{Q}_{S,n}$,$\widehat{Q}_{ST,n}$ and $\widehat{Q}_{T,n}$. We denote the resulting test statistic by $\widehat{LR}1_n$. The analogue of $MLR1$ and $LM1$ are denoted by $\widehat{MLR}1_n$ and $\widehat{LM}1_n$, respectively.

The critical value function for each test statistic ψ is simply $\kappa_{\psi,\alpha}(\widehat{Q}_{T,n})$, as defined in (1.17).

1.6.2 Weak IV Asymptotic Results

In this section we derive the limit distribution of each one-sided test present in the previous section. In particular, we show that all the tests converges in distribution to their respective finite sample distribution under normality.

Under Assumptions WIV-FA and 1-4, Lemma 4 of AMS06a shows that

$$\begin{aligned} &(S_n, T_n) \to_d (S_\infty, T_\infty), \\ &(\widehat{S}_n, \widehat{T}_n, \widehat{\Omega}_n) - (S_n, T_n, \Omega) \to_p 0, \\ &(\widehat{S}_n, \widehat{T}_n, \widehat{\Omega}_n) \to_d (S_\infty, T_\infty, \Omega), \end{aligned} \tag{1.51}$$

where S_∞ and T_∞ are independent k-vectors which are defined as follows:

$$\begin{aligned} vec(N_Z) &\sim N(vec(D_Z C a'), \Omega \otimes D_Z), \\ S_\infty &= D_Z^{-1/2} N_Z b_0 \cdot (b_0' \Omega b_0)^{-1/2} \sim N(c_\beta D_Z^{1/2} C, I_k), \\ T_\infty &= D_Z^{-1/2} N_Z \Omega^{-1} a_0 \cdot (a_0' \Omega^{-1} a_0)^{-1/2} \sim N(d_\beta D_Z^{1/2} C, I_k), \text{ where} \\ D_Z &= D_{11} - D_{12} D_{22}^{-1} D_{21}, \\ D &= \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}, \; D_{11} \in \mathbb{R}^{k \times k}, \; D_{12} \in \mathbb{R}^{k \times p}, \text{ and } D_{22} \in \mathbb{R}^{p \times p}. \end{aligned} \tag{1.52}$$

The matrix D_Z is the probability limit of $n^{-1}Z'Z$. Under $H_0 : \beta = \beta_0$, S_∞ has mean zero, but T_∞ does not. Let

$$\begin{aligned} Q_\infty &= [S_\infty : T_\infty]'[S_\infty : T_\infty], \\ Q_{S,\infty} &= S_\infty' S_\infty, Q_{ST,\infty} = S_\infty' T_\infty, \; Q_{T,\infty} = T_\infty' T_\infty, \\ \mathcal{S}_{2,\infty} &= S_\infty' T_\infty / (||S_\infty|| \cdot ||T_\infty||), \text{ and} \\ \lambda_\infty &= C' D_Z C. \end{aligned} \tag{1.53}$$

By (1.52), we find that the finite-sample distribution of $(Q_{S,\infty}, Q_{ST,\infty}, Q_{T,\infty})$ is the same as that of $(Q_{S,n}, Q_{ST,n}, Q_{T,n})$ with λ_n replaced by λ_∞. Then the asymptotic distribution of the feasible tests and statistics are the same as finite-sample case:

Theorem 4 *Under Assumptions* WIV-FA, 1-4 *and* $\widehat{\underline{k}}$ *given in* 1.50*:*
(a) (i) $(\widehat{t}(\widehat{\underline{k}}), \kappa_{t(\underline{k}),\alpha}(\widehat{Q}_{T,n})) \to_d (t(\underline{k}_\infty), \kappa_{t(\underline{k}),\alpha}(Q_{T,\infty}))$,
(ii) $(\widehat{LM}1_n, z_\alpha) \to_d (LM1(Q_{1,\infty}, Q_{T,\infty}), z_\alpha)$,
(iii) $(\widehat{LR}1_n, \kappa_{LR1,\alpha}(\widehat{Q}_{T,n})) \to_d (LR1(Q_{1,\infty}, Q_{T,\infty}), \kappa_{LR1,\alpha}(Q_{T,\infty}))$,
(iv) $(\widehat{MLR}1_n, \kappa_{MLR1,\alpha}(\widehat{Q}_{T,n})) \to_d (MLR1(Q_{1,\infty}, Q_{T,\infty}), \kappa_{MLR1,\alpha}(Q_{T,\infty}))$.
(b) (i) $P(\widehat{t}(\widehat{\underline{k}}) > \kappa_{t(\underline{k}),\alpha}(\widehat{Q}_{T,n})) \to P(t(\underline{k}_\infty) > \kappa_{t(\underline{k}),\alpha}(Q_{T,\infty}))$,
(ii) $P(\widehat{LM}1_n > z_\alpha) \to P(LM1(Q_{1,\infty}, Q_{T,\infty}) > z_\alpha)$,
(iii) $P(\widehat{LR}1_n > \kappa_{LR1,\alpha}(\widehat{Q}_{T,n})) \to P(LR1(Q_{1,\infty}, Q_{T,\infty}) > \kappa_{LR1,\alpha}(Q_{T,\infty}))$,
(iv) $P(\widehat{MLR}1_n > \kappa_{MLR1,\alpha}(\widehat{Q}_{T,n})) \to P(MLR1(Q_{1,\infty}, Q_{T,\infty}) > \kappa_{MLR1,\alpha}(Q_{T,\infty}))$.
(c)(i) $P(t(\underline{k}_\infty) > \kappa_{t(\underline{k}),\alpha}(Q_{T,\infty})) = P(LM1(Q_{1,\infty}, Q_{T,\infty}) > z_\alpha) = \alpha$ *for 2SLS, LIML and Fuller estimators when* $\beta = \beta_0$*,*
(ii) $P(t(\underline{k}_\infty) > \kappa_{t(\underline{k}),\alpha}(Q_{T,\infty})) \leq \alpha$ *for B2SLS estimator when* $\beta = \beta_0$*,*
(iii)$P(LR1(Q_{1,\infty}, Q_{T,\infty}) > \kappa_{LR1,\alpha}(Q_{T,\infty})) = P(MLR1(Q_{1,\infty}, Q_{T,\infty}) > \kappa_{MLR1,\alpha}(Q_{T,\infty})) = \alpha$ *when* $\beta = \beta_0$ *and provided the significance level* $\alpha \in (0, 1/2)$

Comments. 1. The t-statistics $t(\underline{k}_{\,\infty})$ are the limiting distribution of the t-statistics $\widehat{t}(\widehat{\underline{k}})$.

2. Part (c) asserts that the conditional tests derived from the $\widehat{LM1}_n$ statistic and from the t-statistics based in either the 2SLS, LIML or Fuller estimators are asymptotically similar at level α.

3. The conditional B2SLS t-test is not asymptotically similar when instruments are weak because we set the term $[y_2' P_Z y_2 - (k-2)\omega_{22}]^{1/2}$ to zero if the term inside the square root is negative. However, this test has correct size when $\beta = \beta_0$ by the definition of the critical value function.

1.7 Strong IV Asymptotics

Two important large sample properties of tests are *consistency*: the power of the test under the alternative goes to one as the sample size increases; and *asymptotic efficiency* (AE): the test uniformly maximize the asymptotic power among the asymptotically unbiased tests (see Lehmann and Romano (2005)). Given that, we analyze the asymptotic properties of the conditional tests based on $\widehat{LM1}_n$, $\widehat{LR1}_n$, $\widehat{MLR1}_n$, and $\widehat{t}(\widehat{k})_n$ statistics for both local alternatives (SIV-LA) and fixed alternatives (SIV-FA). Under SIV-LA, we establish AE for the one-sided tests and under SIV-FA, we address consistency.

Before analyzing the asymptotic properties of the conditional one-sided tests, we need to establish the limit behaviour of the critical value functions under SIV-LA and SIV-FA. Under both asymptotics, SIV-LA and SIV-FA, Q_T diverges in probability to ∞. We provide the following preliminary results when this occurs.

Lemma 5 *Let z_α be the $1-\alpha$ quantile of the standard normal distribution. As $q_T \to_p \infty$,*
(a) *for any $\underline{k}$ in* (1.19)*, $\kappa_{t(\underline{k}\,),\alpha}(q_T) \to z_\alpha$,*
(b) *for $\alpha \in (0, 1/2)$, $\kappa_{LR1^{1/2},\alpha}(q_T) \to z_\alpha$.*
(c) *for $\alpha \in (0, 1/2)$, $\kappa_{MLR1^{1/2},\alpha}(q_T) \to z_\alpha$.*

Comment. As the sample size increases the statistic $Q_{T,n}$ diverges to infinity in probability which implies, by the previous result, that the critical value functions of the tests considered converge in probability to the $1-\alpha$ quantile of the standard normal distribution.

1.7.1 Local Alternatives

For local alternatives, β is local to the null value β_0 as $n \to \infty$.

Assumption SIV-LA. (a) $\beta = \beta_0 + B/n^{1/2}$ for some constant $B > 0$.
(b) π is a fixed non-zero k-vector for all $n \geq 1$.
(c) k is a fixed positive integer that does not depend on n.

We use Lemma 6 of AMS06a to establish the strong IV-local alternative limiting distribution of tests. Under Assumptions SIV-LA and 1-4:

$$
\begin{aligned}
&(S_n, T_n/n^{1/2}) \to_d (S_{B\infty}, \alpha_T), \\
&(\widehat{S}_n, \widehat{T}_n/n^{1/2}, \widehat{\Omega}_n) - (S_n, T_n/n^{1/2}, \Omega) \to_p 0, \\
&(\widehat{S}_n, \widehat{T}_n/n^{1/2}, \widehat{\Omega}_n) \to_d (S_{B\infty}, \alpha_T, \Omega),
\end{aligned} \tag{1.54}
$$

where $S_{B\infty}$ and α_T are k-vectors defined as follows:

$$
\begin{aligned}
S_{B\infty} &\sim N(\alpha_S, I_k), \text{ where} \\
\alpha_S &= D_Z^{1/2} \pi B (b_0' \Omega b_0)^{-1/2} \text{ and} \\
\alpha_T &= D_Z^{1/2} \pi (a_0' \Omega^{-1} a_0)^{1/2}.
\end{aligned} \tag{1.55}
$$

These definitions allow us to determine the behavior of the $\widehat{LM1}_n$, $\widehat{LR1}_n$, $\widehat{MLR1}_n$, and $\widehat{t}(\widehat{k})_n$ statistics under SIV-LA asymptotics.

Theorem 6 *Under Assumptions* SIV-LA *and* 1-4*:*
(a) *if* $\widehat{\underline{k}} = \underline{k} + O_p(n^{-1}) = 1 + O_p(n^{-1})$, *then* $\widehat{t}(\widehat{\underline{k}}) = t\left(\underline{k}\right) + o_p(1) \to_d (\alpha_T' S_{B\infty})/||\alpha_T||$.
(b) $\widehat{LM1}_n = LM1_n + o_p(1) \to_d (\alpha_T' S_{B\infty})/||\alpha_T||$,
(c) $\widehat{LR1}_n^{1/2} = LR1_n^{1/2} + o_p(1) \to_d \max\left\{(\alpha_T' S_{B\infty})/||\alpha_T||, 0\right\}$,
(d) $\widehat{MLR1}_n^{1/2} = MLR1_n^{1/2} + o_p(1) \to_d \max\left\{(\alpha_T' S_{B\infty})/||\alpha_T||, 0\right\}$,

Comments. 1. The requirement $\underline{k} = 1 + O_p\left(n^{-1}\right)$ is satisfied by the $\underline{k}$-class estimators considered.

2. The requirement $\widehat{\underline{k}} = \underline{k} + O_p(n^{-1})$ allows us to show that the replacement of $\underline{k}$ by $\widehat{\underline{k}}$ does not have any asymptotic effect for the t-statistics.

Together with Lemma 5, Theorem 6 yields the following optimality result for a sequence of experiments under SIV-LA and i.i.d normal errors with unknown covariance matrix Ω. Under SIV-LA, the curvature of the model (1.7) vanishes asymptotically and standard local asymptotically normal (LAN) likelihood ratio theory is applicable. For one-sided alternatives, the usual one-sided LM test is AE under the SIV-LA asymptotics and i.i.d normal errors. The others one-sided tests that we propose are also AE.

Theorem 7 *Suppose Assumptions SIV-LA and* 1 *hold and the reduced-form errors* $\{V_i : i \geq 1\}$ *are i.i.d normal, independent of* $\{\overline{Z}_i : i \geq 1\}$ *with mean zero and p.d variance matrix* Ω *which may be known or unknown. Then the score test based on* $\widehat{LM1}_n$ *and the conditional tests based on* $\widehat{t}\left(\underline{k}\right)_n$ *are one-sided AE. If* $\alpha \in (0, 1/2)$, *the conditional tests based on* $\widehat{LR1}_n^{1/2}$ *and* $\widehat{MLR1}_n^{1/2}$ *are also AE.*

1.7.2 Fixed Alternatives

We now analyze properties of the tests under strong IV fixed alternative (SIV-FA) asymptotics. This asymptotic framework is novel in the weak-instrument literature and determines the *consistency* of tests.

Assumption SIV-FA. (a) $\beta \neq \beta_0$ is a fixed scalar for all $n \geq 1$.
(b) π is a fixed non-zero k-vector for all $n \geq 1$.
(c) k is a fixed positive integer that does not depend on n.

The strong IV-fixed alternative (SIV-FA) asymptotic behavior of tests depends on the random vector $\varsigma_k \sim N(0, I_k)$ and

$$\lambda_{FA} = \pi' D_Z \pi, \tag{1.56}$$

where D_Z is defined in (1.52).

Lemma 8 *Under Assumptions* SIV-FA *and* 1-3,
(i) $(S_n/n^{1/2}, T_n/n^{1/2}) \to_p (c_\beta D_Z^{1/2}\pi,\ d_\beta D_Z^{1/2}\pi)$,
(ii) $(\widehat{S}_n/n^{1/2}, \widehat{T}_n/n^{1/2}) - (S_n/n^{1/2}, T_n/n^{1/2}) \to_p 0$, *and*
(iii) *if* $\beta = \beta_{AR}$ *and Assumption* 4 *holds, then* $T_n \to_d \varsigma_k$ *and* $\widehat{T}_n - T_n \to_p 0$.

Lemma 8 allows us to determine the limiting behavior of the one-sided conditional tests based on $\widehat{t}\left(\widehat{\underline{k}}\right)_n$, $\widehat{LR1}_n$, $\widehat{MLR1}_n$ and $\widehat{LM1}_n$ statistics under SIV-FA asymptotics.

Theorem 9 *Under Assumptions SIV-FA and* 1-3,
(g) *if* $\widehat{\underline{k}} = \underline{k} + o_p(1) = 1 + o_p(1)$, $\widehat{t}(\widehat{\underline{k}})/n^{1/2} = t\left(\underline{k}\right)/n^{1/2} + o_p(1) \to_p c_\beta \lambda_{FA}^{1/2} \times (b_0'\Omega b_0/b'\Omega b)^{1/2}$.
(b) *if* $\beta \neq \beta_{AR}$, *then* $\widehat{LM1}_n/n^{1/2} = LM1_n/n^{1/2} + o_p(1) \to_p c_\beta \lambda_{FA}^{1/2}$,
(c) *if* $\beta = \beta_{AR}$ *and Assumption* 4 *holds, then* $\widehat{LM1}_n/n^{1/2} = LM1_n/n^{1/2} + o_p(1) \to_d c_\beta \pi' D_Z^{1/2} \varsigma_k/||\varsigma_k||$,
(d) *if* $\beta > \beta_0$, $\widehat{LR1}_n^{1/2}/n^{1/2} = LR1_n^{1/2}/n^{1/2} + o_p(1) \to_p c_\beta \lambda_{FA}^{1/2}$, *and*
(e) *if* $\beta < \beta_0$, $\widehat{LR1}_n^{1/2}/n^{1/2} = LR1_n^{1/2}/n^{1/2} + o_p(1) \to_p \sqrt{max\left(c_\beta^2 - \omega_{22}^{-1}, 0\right)} \lambda_{FA}^{1/2}$,
(f) *if* $\beta > \beta_0$, $\widehat{MLR1}_n^{1/2}/n^{1/2} = MLR1_n^{1/2}/n^{1/2} + o_p(1) \to_p \sqrt{min\left(c_\beta^2, \omega_{22}^{-1}\right)} \lambda_{FA}^{1/2}$,

(g) *if* $\beta < \beta_0$, $\widehat{MLR1}_n^{1/2}/n^{1/2} = MLR1_n^{1/2}/n^{1/2} + o_p(1) \to_p 0$,

Comments. 1. When $\beta \neq \beta_{AR}$, the critical values of the conditional tests are either constants or converge in probability to constants as $n \to \infty$ (see the comments following Lemma 5). When $\beta = \beta_{AR}$, the critical value functions of these tests (for each β_0) are bounded. Therefore, this theorem addresses the consistency of each test.

2. The one-sided LM test rejects the null when $\widehat{LM1}_n/n^{1/2} > z_\alpha/n^{1/2}$. Because $z_\alpha/n^{1/2}$ converges to zero and the probability of $c_\beta \pi' D_Z^{1/2} \varsigma_k/||\varsigma_k||$ being smaller than zero equals 50%, the LM1 test is *not* consistent at $\beta = \beta_{AR}$.

3. Part (d) shows that the $CLR1$ test is consistent against any alternative $\beta > \beta_0$ for the null hypothesis $H_0 : \beta = \beta_0$. Part (d) shows that the $CLR1$ test asymptotically rejects the null with probability one for some value of $\beta < \beta_0$. Hence, the $CLR1$ test has asymptotic size equal to one once we augment the null hypothesis to $H_0 : \beta \leq \beta_0$.

4. Parts (f)-(g) establish consistency of the conditional tests based on the $MLR1$ statistic whether $H_0 : \beta = \beta_0$ or $H_0 : \beta \leq \beta_0$.

5. 6. Consider the ad-hoc statistic $ALR1 = LR \times 1(\beta(\underline{k}_{LIMLK}) \geq \beta_0)$ instead of the $LR1$ and $MLR1$ statistics. Under SIV-LA, the $CALR1$ test is AE. Under SIV-FA, $ALR1^{1/2}/n^{1/2}$ converges to $c_\beta \lambda_{FA}^{1/2}$ when $\beta > \beta_0$ and to zero when $\beta < \beta_0$. Because $c_\beta \lambda_{FA}^{1/2} \geq \sqrt{\min\{c_\beta^2, w_{22}^{-1}\}}\lambda_{FA}^{1/2}$ found in part (f), the $CALR1$ test could dominate $CMLR1$ under strong instruments. For example, if $\beta_0 = 0$ and $w_{11} = w_{22}$ then $c_\beta \lambda_{FA}^{1/2} \geq \sqrt{\min\{c_\beta^2, w_{22}^{-1}\}}\lambda_{FA}^{1/2}$ holds when $\beta > 1$. This finding suggests comparing $ALR1$ and $MLR1$ statistics under either Bahadur or Hodges-Lehmann efficiency instead of Pitman drifts (i.e., SIV-LA asymptotics). We leave this theoretical exercise for future research.

1.8 Numerical Results

The numerical simulations apply asymptotically to feasible tests which replace Ω with $\widehat{\Omega}$ for stochastic regressors and non-normal errors. Following Section 6.4 of AMS06a, the power envelopes obtained here are asymptotically valid when the errors are iid normal with *unknown* covariance matrix. Numerical simulations have been computed at significance level $\alpha = 0.05$ for $\lambda/k = 0.5, 1, 2, 4, 8, 16$, which span the range from weak to strong instruments, $\rho = 0.2$, 0.5, 0.9, and $k = 2$, 5, 10, 20. To conserve space, we focus here on testing $H_0 : \beta \leq 0$ against $H_1 : \beta > 0$ when $\lambda/k = 1, 2$, $\rho = 0.5, 0.9$, and $k = 5$. Additional numerical simulations are available in the supplement (including testing $H_0 : \beta \geq 0$ against $H_1 : \beta < 0$).

The simulations are presented as plots of power envelopes and power functions against various alternative values of β and λ. Power is plotted as a function of the rescaled alternative $\beta\lambda^{1/2}$. This can be thought of as a local power plot, where the local neighborhood is $\lambda^{1/2}$ instead of the usual $1/n^{1/2}$, since λ measures the effective sample size. We report simulations for all four conditional t-tests, the CLR1, and the CMLR1 test.

Figures 1.1 to 1.4 assess the power properties of several tests for $\rho = 0.5$ and $\rho = 0.9$. We report power curves for the conditional t-tests as well as both CLR1 and CMLR1 tests. Conditional critical values for all test statistics are computed based on 100,000 Monte Carlo simulations for each observed value $Q_T = q_T$. In the absence of a UMPI test, we consider tests whose power functions may be near the one-sided power envelope for invariant similar tests based on Corollary 2. In the supplement, we provide numerical evidence that the power envelopes for similar and non-similar tests are alike.

The CLR1 test has rejection probabilities close to the power upper bound for alternatives $\beta > \beta_0$. However, this test has null rejection probabilities close to one for small enough values of $\beta < \beta_0$. This poor behavior is in accordance with Theorem 9 which shows that the CLR1 test is not consistent. Hence, this test is not very useful for applied researchers[8].

The CMLR1 and all one-sided conditional t-tests do have correct size for $H_0 : \beta \leq \beta_0$. Perhaps surprisingly, the conditional t-tests based on the 2SLS and Fuller estimator have good performance. The conditional

[8] Additional numerical results show that the $CALR1$ test based on the $ALR1 = LR \times 1(\beta(\underline{k}_{LIMLK}) \geq \beta_0)$ statistic also does not have correct size for testing $H_0 : \beta \leq 0$. Its null rejection probability can be close to 20% when $\lambda/k = 0.5$ and to 10% when $\lambda/k = 1$ for values of β away from zero.

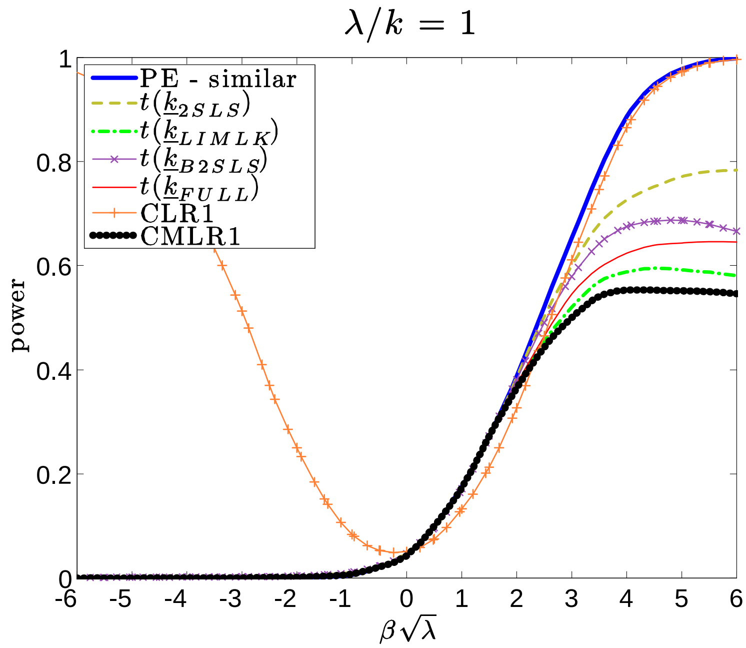

Figure 1.1: Asymptotic power of one-sided conditional tests ($\rho = 0.5$ and $\lambda/k = 1$).

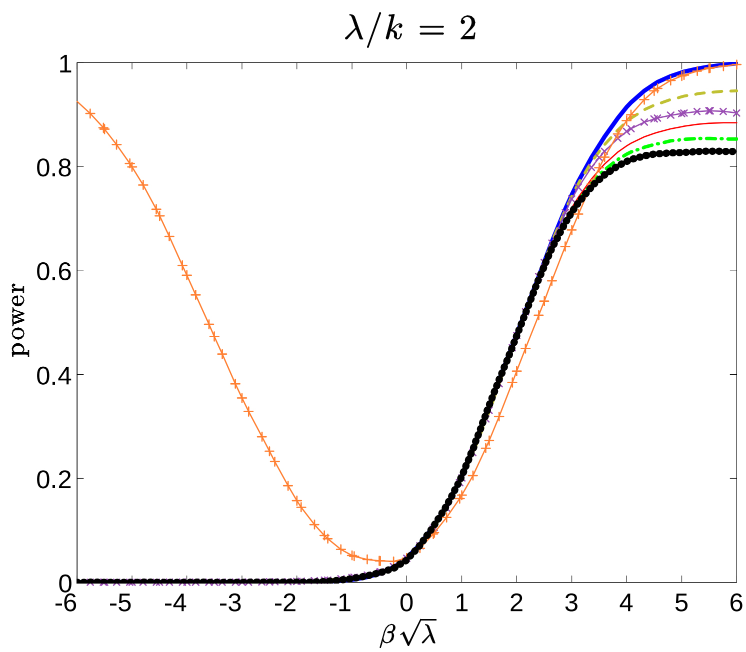

Figure 1.2: Asymptotic power of one-sided conditional tests ($\rho = 0.5$ and $\lambda/k = 2$).

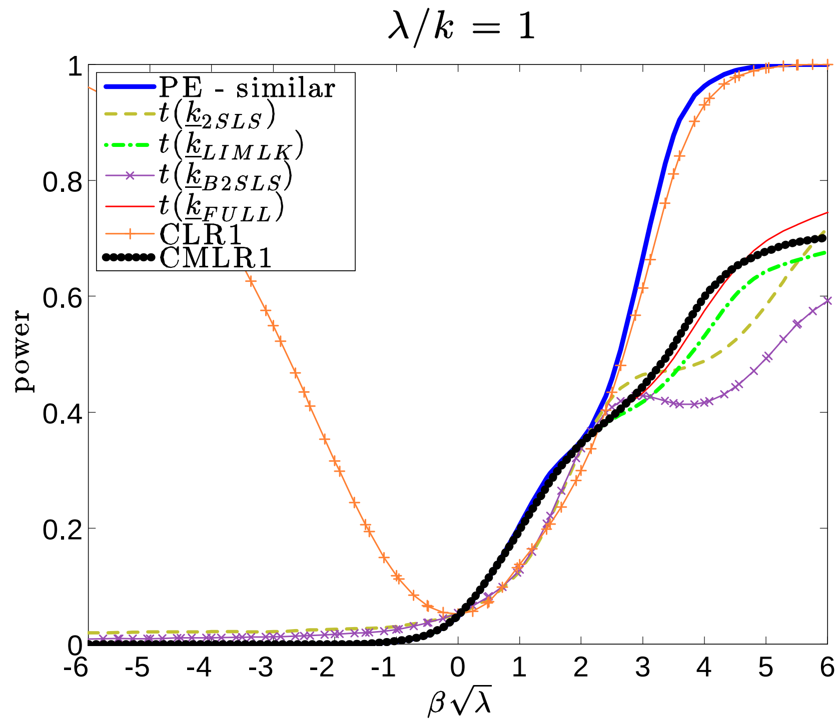

Figure 1.3: Asymptotic power of one-sided conditional tests ($\rho = 0.9$ and $\lambda/k = 1$).

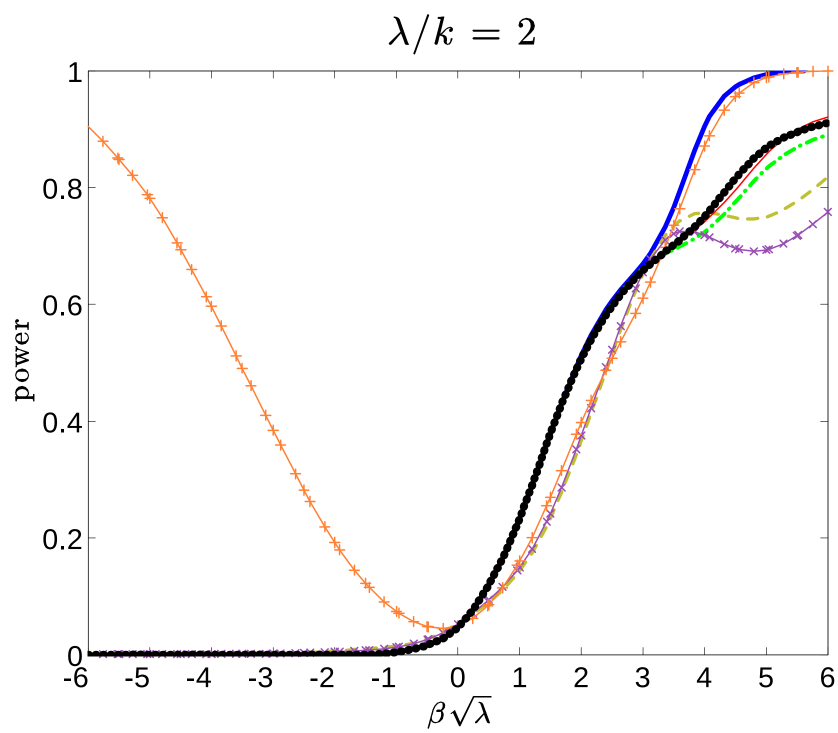

Figure 1.4: Asymptotic power of one-sided conditional tests ($\rho = 0.9$ and $\lambda/k = 4$).

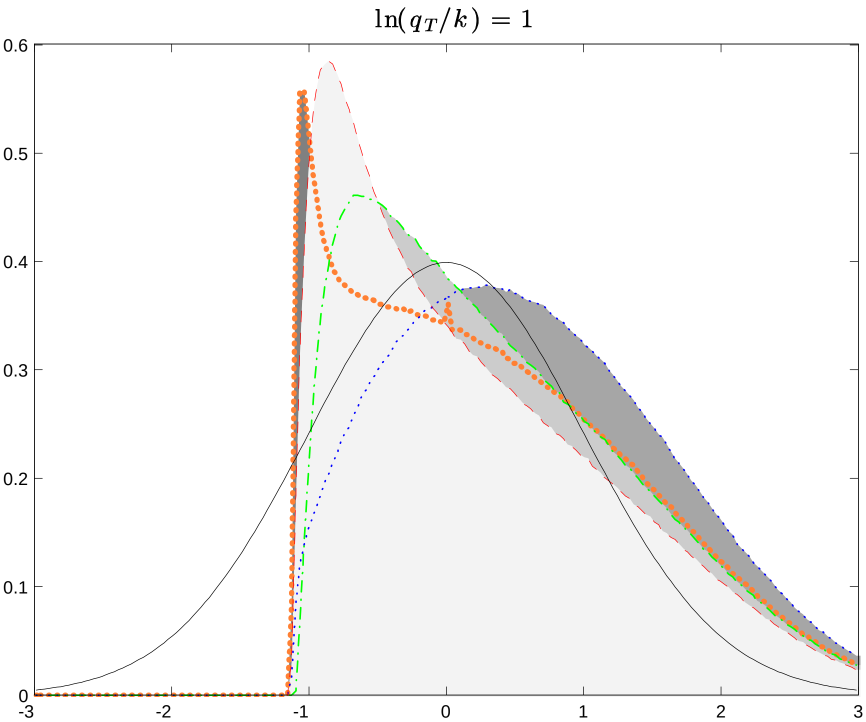

Figure 1.5: Probability density function for $t(\underline{k})$ conditional on Q_T, where $\ln(q_T/k) = 1$.

test based on the 2SLS estimator numerically outperforms the one based on the B2SLS estimator. The test based on the Fuller estimator dominates the LIML counterpart and the CMLR1 test.

As λ increases, the power of the conditional t-tests approaches the conservative power envelope. This result is in accordance with subsection 1.7.1, which shows that the conditional t-tests (along with the CLR1 and CMLR1 tests) are asymptotically efficient under strong-instrument asymptotics. When λ/k is as small as 2, the conditional 2SLS t-test performs near the power envelope for $\rho = 0.5$ while the conditional Fuller t-test has power close to the power envelope for $\rho = 0.9$. The supplement provides further evidence for the use of the one-sided conditional t-tests (in particular, those based on the 2SLS and Fuller estimators) in empirical practice.

1.9 Two-Sided Tests

In this section we are interested in revise the finding of AMS07 for the conditional t-tests in the two-sided hypothesis:

$$H_0 : \beta = \beta_0 \text{ against } H_1 : \beta \neq \beta_0.$$

In section 1.8 and in the supplements we presented numerical results showing a good performance for the one-sided conditional t-tests, this is striking given the poor performance of the two-sided conditional t-tests documented by AMS07. The goal of this section is to solve this apparent counterintuitive result between one-sided and two-sided conditional t-tests.

It turns out that AMS07's finding strongly relies on the asymmetry of the null conditional distribution of the t-statistics considered. Figure 1.5 plots the standard normal distribution with the distributions for the $t(\underline{k})$ statistics based on the 2SLS, LIMLK, Fuller and B2SLS to contrast this asymmetry. When the instruments are weak (q_T is small), the null conditional distribution of the t-statistics are asymmetric around zero and the use of a decision rule symmetric in the critical value function generates biased tests. As the strength of the instruments increases (q_T grows), the null conditional distribution of the t-statistics become symmetric and we can use the symmetric critical value functions to reject the null hypothesis.To overcome this asymmetry we propose two methods: the first method augment the conditional argument of Moreira

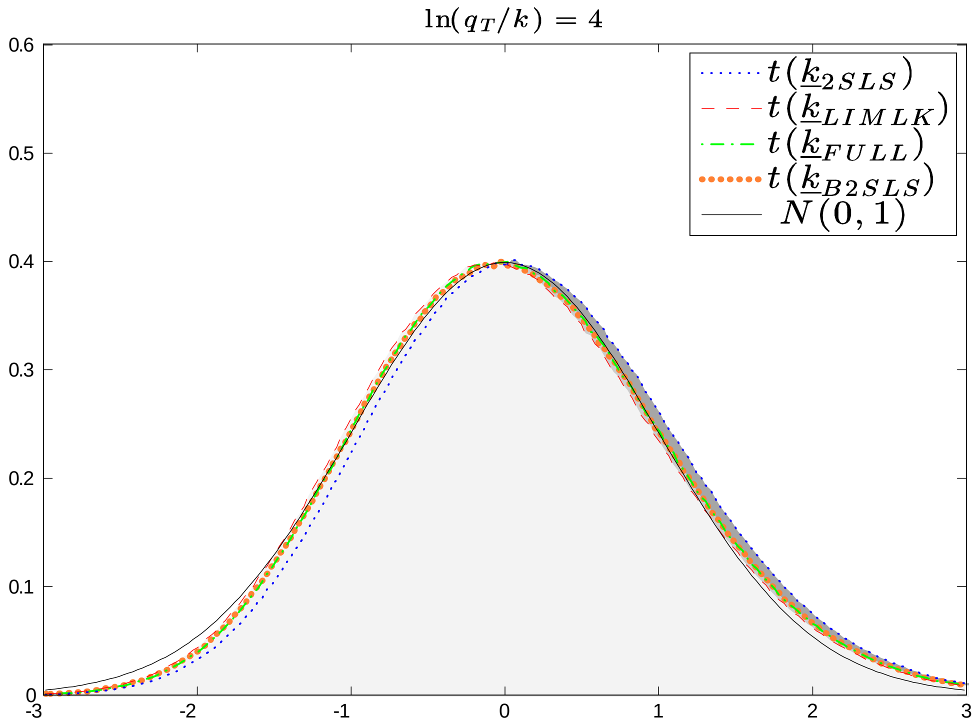

Figure 1.6: Probability density function for $t(\underline{k})$ conditional on Q_T, where $\ln(q_T/k) = 4$.

(2003) in a manner to obtain two critical value functions; the second method use other t-statistics, which are approximately symmetric, in the construction of the t-tests.

Hereinafter, it is convenient to work with the statistics

$$Q_{(k-1)} = S'M_T S,\ LM1 = Q_{ST}/Q_T^{1/2},\text{ and } Q_T, \tag{1.57}$$

which are a one-to-one transformation of Q.

For testing $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$, Theorem 1 of AMS06b proves that an unbiased test ϕ must satisfy

$$E_{\beta_0}\left(\phi\left(Q_{(k-1)}, LM1, q_T\right)\right) = \alpha \text{ and} \tag{1.58}$$

$$E_{\beta_0}\left(\phi\left(Q_{(k-1)}, LM1, q_T\right) \cdot LM1\right) = 0 \tag{1.59}$$

for almost all values of q_T. By Corollary 1 of AMS06b, the CLR test satisfies both boundary conditions. Other conditional tests – such as tests based on $t\left(\underline{k}\right)^2$– do not necessarily satisfy (1.59). This places considerable limits on the applicability of conditional method of generating unbiased tests.

For this consider the two-sided unbiased tests based on one-sided statistics $t\left(\underline{k}\right)$ which rejects the null when

$$t\left(\underline{k}\right) < \kappa_{t(\underline{k}),1-x_\alpha}\left(q_T\right) \text{ or } t\left(\underline{k}\right) > \kappa_{t(\underline{k}),\alpha-x_\alpha}\left(q_T\right), \tag{1.60}$$

where $\kappa_{t(\underline{k}),x_\alpha}\left(q_T\right)$ is the $1-x_\alpha$ quantile of the conditional distribution and $x_\alpha \in [0, \alpha]$ is chosen to approximately satisfy (1.58) and (1.59). Inverting the approximately unbiased t-tests in (1.60) allows us to construct confidence regions around a chosen estimator (we do not obtain equal-tailed two-sided intervals, otherwise the test would be biased). In particular, we can construct confidence regions based on the 2SLS estimator, which is commonly used in applied research.

Implicit the conditional t-test used in AMS07 consider a symmetric null distribution of $t\left(\underline{k}\right)$ conditional on q_T. Given that and $x_\alpha = \alpha/2$ we have

$$\kappa_{t(\underline{k}),1-x_\alpha}\left(q_T\right) = -\kappa_{t(\underline{k}),\alpha-x_\alpha}\left(q_T\right), \tag{1.61}$$

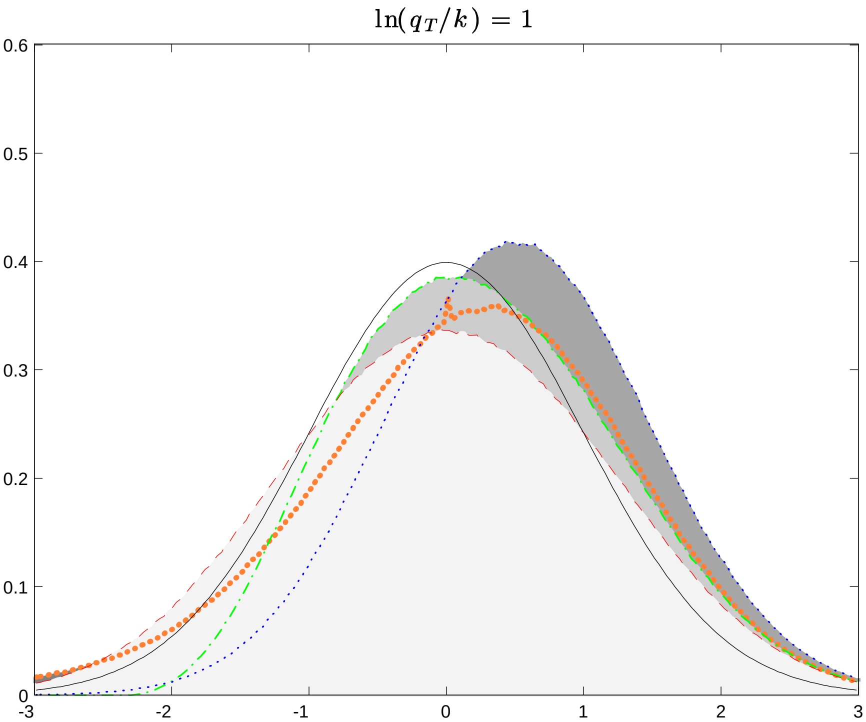

Figure 1.7: Probability density function for $t_0(\underline{k})$ conditional on Q_T, where $\ln(q_T/k) = 1$.

Consequently, the test that reject the null when (1.60) is equivalent as the test that rejects the null when $t\left(\underline{k}\right)^2 > \left(\kappa_{t(\underline{k}),\alpha/2}\left(q_T\right)\right)^2$. That is, we would have obtained the conditional test based on $t\left(\underline{k}\right)^2$ where the critical value function is $\kappa_{t(\underline{k})^2,\alpha}\left(q_T\right) = \left(\kappa_{t(\underline{k}),\alpha/2}\left(q_T\right)\right)^2$.

The second method considered use t-tests based on modifications of the t-statistic that are approximately symmetric. Figures 1.7 and 1.8 plot the conditional distribution for the modified versions of t-statistics:

$$t_0\left(\underline{k}\right) = \frac{\beta\left(\underline{k}\right) - \beta_0}{\sigma_0 \cdot \left[y_2' P_Z\, y_2 - n\left(\underline{k} - 1\right)\omega_{22}\right]^{-1/2}}, \tag{1.62}$$

where $\sigma_0^2 = (1, -\beta_0)'\Omega(1, -\beta_0)$, estimator of the variance of structural error. The conditional distributions for the $t_0\left(\underline{k}\right)$ statistics based on the 2SLS and Fuller estimators are also asymmetric around zero when q_T is small. However, the $t_0\left(\underline{k}\right)$ statistic for the LIMLK estimator is nearly symmetric around zero for any value of q_T. Hence, the conditional test based on $t_0\left(\underline{k}_{\,LIMLK}\right)^2$ is nearly unbiased and should not suffer the poor power properties found by AMS07 for the $t\left(\underline{k}\right)^2$ statistics.

In the supplement, we provide numerical results showing that the conditional t-test based on the $t_0\left(\underline{k}_{\,LIMLK}\right)^2$ statistic and some of the unbiased t-tests can perform as well as the CLR test. Hence, the conclusion of AMS07 is only valid for a smaller class of t-tests.

1.10 Uniform Convergence

In this section we present a proposition showing that the one-sided conditional t-tests based on 2SLS, LIML and Fuller estimator are asymptotic similar, in a uniform sense. The proof uses Corollary 2.1 of Andrews, Cheng, and Guggenberger (2011), where we check convergence of the null rejection probability under all parameter subsequences. We suppose the model given by equations (1.1) and (1.2) satisfies:

Assumption S1. $\{(X_i', \widetilde{Z}_i', u_i, v_{2i})' : i \in \{1, ..., n\}\}$ are i.i.d. with distribution F,
Assumption S2. $E_F(V_i \otimes \overline{Z}_i) = 0$ where $V_i = (u_i, v_{2i})$ and $\overline{Z}_i = [\widetilde{Z}_i' : X_i']$
Assumption S3. $E_F(V_i V_i' | \overline{Z}_i) = E_F(V_i V_i')$.

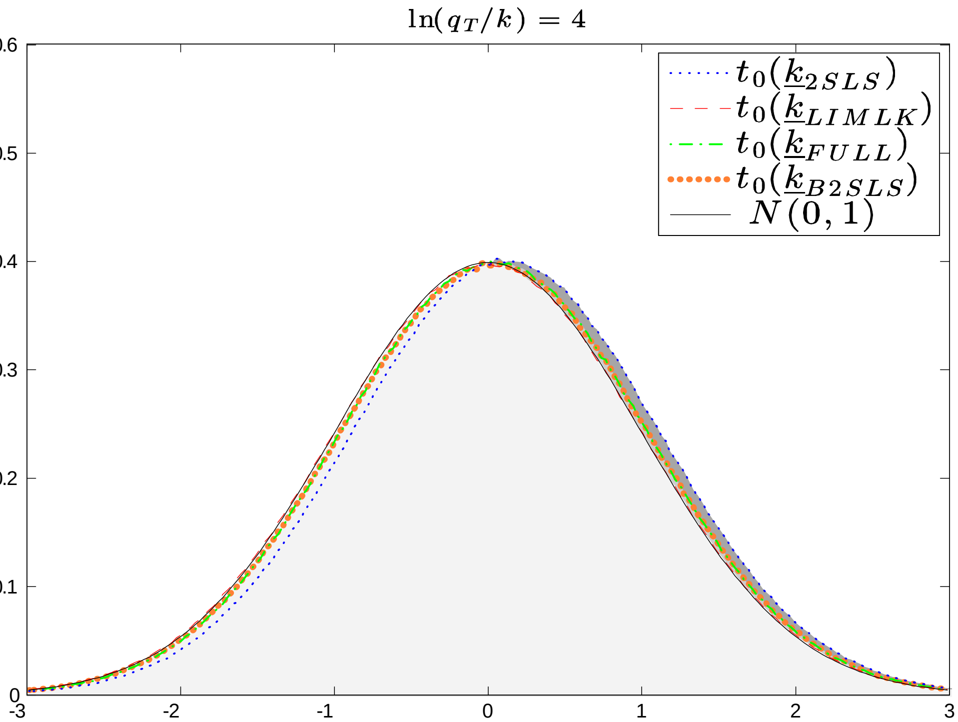

Figure 1.8: Probability density function for $t_0(\underline{k})$ conditional on Q_T, where $\ln(q_T/k) = 1$.

Proposition 10 *The one-sided conditional t-tests based on 2SLS, LIML and Fuller estimator are asymptotic similar in a uniform sense.*

1.11 Empirical Example

We revisit Angrist and Krueger's (1991) work on returns to schooling for the 1920-29, 1930-39, and 1940-49 cohorts using the U.S. Census. The data consist of the original 1970 and 1980 U.S. Census sample of males born in the United States used by Angrist and Krueger (1991)[9]. The 1920-29 cohort is from the 1970 Census sample, the 1930-39 and 1940-49 cohorts are both from the 1980 Census sample. For the 1920-29 cohort we excluded 3000 individuals whose place-of-birth was ambiguous, leaving us with a sample size of 244,099 observations. All individuals in the 1930-39 and 1940-49 cohorts had clearly identified places-of-birth and none were eliminated. The sample sizes of the 1930-39 and 1940-49 cohorts are 329,509 and 486,926 observations, respectively.

We include a constant, race, metropolitan area, marital status, age, age-squared, and dummies for year-of-birth, state-of-birth, and regions as covariates. The log-weekly earning variable is imputed from weeks worked and annual earnings. The metropolitan area variable equals one if the individual lives in a Standard Metropolitan Statistical Area (SMSA), as determined by the U.S. Office of Management and Budget and recorded by the Census. The 50 state-of-birth dummies are constructed from the place-of-birth Census code variable, ignoring the code value of 11 (District of Columbia). The race variable equals one if the individual is black and zero if not. The 8 regional dummy variables are constructed from the Census region code variable that represents the 9 Regional divisions used by the U.S. Census Bureau. The age variable is imputed from the year-of-birth and quarter-of-birth of individuals. A detailed description of how the data were constructed from the original Census data can be found in Appendix 1 of Angrist and Krueger (1991).

Tables 1 to 3 provide specifications for the empirical example. Specifications I and II use quarter-of-birth and quarter-of-birth $\times$ year-of-birth respectively as instruments, and include a constant, race, metropolitan area and marital status, nine year-of-birth and eight regional dummies as controls. Specification III adds

[9] The data are available at http://economics.mit.edu/faculty/angrist/data1/data.

age and age^2 as covariates and allows interaction between quarter-of-birth and year-of-birth. Specification IV replaces year-of-birth dummies used in specification III by state-of-birth dummies.

The first rows present the OLS, LIML, 2SLS, Fuller and B2SLS estimators with their respective standard errors. We next present 95% confidence confidence intervals for one-sided tests: CMLR1 and unbiased t-tests based on the LIML, 2SLS, Fuller and B2SLS estimators. Finally we present 95% confidence intervals for two-sided tests: CLR and unbiased t-tests based on the LIML, 2SLS, Fuller and B2SLS estimators.

We focus on specification IV, shown by Cruz and Moreira (2005) to be informative for returns to schooling despite the low first-stage F-statistic and 178 instruments. The LIML estimator is larger than the 2SLS estimator for the three cohorts. We first report confidence intervals for one-sided tests: CMLR1 and the t-tests based on the 2SLS and LIML estimators. One-sided tests may be appropriate in this application. Because returns to schooling are non-negative, we can test against the alternative $H_1 : \beta > \beta_0(= 0)$. The confidence regions for the t-tests are comparable despite the associated estimates being quite different. For example, take the 1930-39 cohort. The 2SLS estimates returns to schooling to be 8.1% while the LIML estimate is 9.8%. The lower bound for our confidence regions is about the same (6.1%-6.3%). In the one-sided case we do not make a statement about the upper bound on returns to schooling.

Next we report confidence intervals for two-sided tests: the CLR test and unbiased t-tests based on the 2SLS and LIML estimators. As in the one-sided case, confidence regions for the t-tests and CLR test are comparable. The lower bound for our confidence regions of the 1930-39 cohort is about the same across two-sided tests (5.4-5.6%), while the upper bound for the unbiased t-tests using the 2SLS estimator is slightly larger than using the LIML estimator (15.5% instead of 14.5%). The unbiased t-tests and the CLR test produce comparable confidence intervals. In particular, the confidence regions centered around the 2SLS estimator can be informative even when the first-stage F-statistic is low. This empirical exercise supports our theoretical work on the use of the 2SLS and confidence intervals based on unbiased t-tests in applied work.

Table 1: Effects of Years of Education on Log Weekly Earnings (1920-29 cohort)

	I	II	III	IV
$\widehat{\beta}(\widehat{\underline{k}}_{OLS})$ $(s.e)$	0.0701 (0.004)	0.0701 (0.004)	0.0701 (0.004)	0.0692 (0.004)
$\widehat{\beta}(\widehat{\underline{k}}_{LIML})$ $(s.e)$	0.0595 (0.0166)	0.0691 (0.0151)	0.2615 (0.1116)	0.1390 (0.0196)
$\widehat{\beta}(\widehat{\underline{k}}_{2SLS})$ $(s.e)$	0.0594 (0.0166)	0.0688 (0.0175)	0.1053 (0.0331)	0.0936 (0.0109)
$\widehat{\beta}(\widehat{\underline{k}}_{FULL})$ $(s.e)$	0.0595 (0.0165)	0.0688 (0.0174)	0.2318 (0.1026)	0.1383 (0.0195)
$\widehat{\beta}(\widehat{\underline{k}}_{B2SLS})$ $(s.e)$	0.0594 (0.0166)	0.0689 (0.0170)	0.3640 (0.1821)	0.1365 (0.0191)
	1-sided 95% Confidence Regions			
$CLR1$	$[0.031, \infty)$	$[0.035, \infty)$	$(-\infty, \infty)$	$[0.089, \infty)$
$CMLR1$	$[0.031, \infty)$	$[0.035, \infty)$	$(-\infty, \infty)$	$[0.089, \infty)$
$t_0(\widehat{\underline{k}}_{LIML})$	$[0.031, \infty)$	$[0.035, \infty)$	$(-\infty, \infty)$	$[0.089, \infty)$
$\widehat{t}(\widehat{\underline{k}}_{LIML})$	$[0.031, \infty)$	$[0.035, \infty)$	$(-\infty, \infty)$	$[0.089, \infty)$
$t_0(\widehat{\underline{k}}_{2SLS})$	$[0.030, \infty)$	$[0.036, \infty)$	$(-\infty, \infty)$	$[0.087, \infty)$
$\widehat{t}(\widehat{\underline{k}}_{2SLS})$	$[0.030, \infty)$	$[0.036, \infty)$	$(-\infty, \infty)$	$[0.087, \infty)$
$t_0(\widehat{\underline{k}}_{FULL})$	$[0.031, \infty)$	$[0.035, \infty)$	$(-\infty, \infty)$	$[0.089, \infty)$
$\widehat{t}(\widehat{\underline{k}}_{FULL})$	$[0.031, \infty)$	$[0.035, \infty)$	$(-\infty, \infty)$	$[0.089, \infty)$
$t_0(\widehat{\underline{k}}_{B2SLS})$	$[0.030, \infty)$	$[0.036, \infty)$	$(-\infty, \infty)$	$[0.087, \infty)$
$\widehat{t}(\widehat{\underline{k}}_{B2SLS})$	$[0.030, \infty)$	$[0.040, \infty)$	$(-\infty, \infty)$	$[0.106, \infty)$
	2-sided 95% Confidence Regions			
CLR	$[0.025, 0.093]$	$[0.028, 0.109]$	$(-\infty, \infty)$	$[0.080, 0.214]$
$t_0(\widehat{\underline{k}}_{LIML})$	$[0.025, 0.093]$	$[0.028, 0.109]$	$(-\infty, \infty)$	$[0.080, 0.213]$
$\widehat{t}(\widehat{\underline{k}}_{LIML})$	$[0.025, 0.093]$	$[0.028, 0.109]$	$(-\infty, \infty)$	$[0.080, 0.213]$
$t_0(\widehat{\underline{k}}_{2SLS})$	$[0.025, 0.093]$	$[0.029, 0.107]$	$(-\infty, \infty)$	$[0.077, 0.236]$
$\widehat{t}(\widehat{\underline{k}}_{2SLS})$	$[0.025, 0.093]$	$[0.029, 0.107]$	$(-\infty, \infty)$	$[0.077, 0.236]$
$t_0(\widehat{\underline{k}}_{FULL})$	$[0.025, 0.093]$	$[0.028, 0.109]$	$(-\infty, \infty)$	$[0.080, 0.213]$
$\widehat{t}(\widehat{\underline{k}}_{FULL})$	$[0.025, 0.093]$	$[0.028, 0.109]$	$(-\infty, \infty)$	$[0.080, 0.213]$
$t_0(\widehat{\underline{k}}_{B2SLS})$	$[0.025, 0.093]$	$[0.029, 0.108]$	$(-\infty, \infty)$	$[0.101, 0.228]$
$\widehat{t}(\widehat{\underline{k}}_{B2SLS})$	$[0.025, 0.093]$	$[0.034, 0.103]$	$(-\infty, \infty)$	$[0.077, 0.180]$
F(first stage)	37.971	4.538	1.055	1.553
k	3	30	28	178

Table 2: Effects of Years of Education on Log Weekly Earnings (1930-39 cohort)

	I	II	III	IV
$\widehat{\beta}(\widehat{\underline{k}}_{OLS})$ $_{(s.e)}$	0.0632 (0.003)	0.0632 (0.003)	0.0632 (0.003)	0.0628 (0.003)
$\widehat{\beta}(\widehat{\underline{k}}_{LIML})$ $_{(s.e)}$	0.0999 (0.0210)	0.0838 (0.0179)	0.0574 (0.0385)	0.0982 (0.0153)
$\widehat{\beta}(\widehat{\underline{k}}_{2SLS})$ $_{(s.e)}$	0.0990 (0.0207)	0.0806 (0.0164)	0.0600 (0.0290)	0.0811 (0.0109)
$\widehat{\beta}(\widehat{\underline{k}}_{FULL})$ $_{(s.e)}$	0.0995 (0.0209)	0.0836 (0.0178)	0.0577 (0.0378)	0.0980 (0.0153)
$\widehat{\beta}(\widehat{\underline{k}}_{B2SLS})$ $_{(s.e)}$	0.0994 (0.0208)	0.0848 (0.0183)	0.0555 (0.0445)	0.1016 (0.0161)
	1-sided 95% Confidence Regions			
$CLR1$	$[0.065, \infty)$	$[0.051, \infty)$	$[-0.184, \infty)$	$[0.061, \infty)$
$CMLR1$	$[0.065, \infty)$	$[0.051, \infty)$	$[-0.184, \infty)$	$[0.061, \infty)$
$t_0(\widehat{\underline{k}}_{LIML})$	$[0.065, \infty)$	$[0.051, \infty)$	$[-0.184, \infty)$	$[0.061, \infty)$
$\widehat{t}(\widehat{\underline{k}}_{LIML})$	$[0.065, \infty)$	$[0.051, \infty)$	$[-0.075, \infty)$	$[0.062, \infty)$
$t_0(\widehat{\underline{k}}_{2SLS})$	$[0.065, \infty)$	$[0.051, \infty)$	$[-0.378, \infty)$	$[0.063, \infty)$
$\widehat{t}(\widehat{\underline{k}}_{2SLS})$	$[0.065, \infty)$	$[0.051, \infty)$	$[-0.270, \infty)$	$[0.063, \infty)$
$t_0(\widehat{\underline{k}}_{FULL})$	$[0.065, \infty)$	$[0.051, \infty)$	$[-0.105, \infty)$	$[0.061, \infty)$
$\widehat{t}(\widehat{\underline{k}}_{FULL})$	$[0.065, \infty)$	$[0.051, \infty)$	$[-0.076, \infty)$	$[0.062, \infty)$
$t_0(\widehat{\underline{k}}_{B2SLS})$	$[0.065, \infty)$	$[0.051, \infty)$	$[-0.946, \infty)$	$[0.063, \infty)$
$\widehat{t}(\widehat{\underline{k}}_{B2SLS})$	$[0.065, \infty)$	$[0.054, \infty)$	$[-0.384, \infty)$	$[0.075, \infty)$
	2-sided 95% Confidence Regions			
CLR	$[0.059, 0.145]$	$[0.044, 0.125]$	$[-0.187, 0.285]$	$[0.055, 0.145]$
$t_0(\widehat{\underline{k}}_{LIML})$	$[0.059, 0.145]$	$[0.044, 0.125]$	$[-0.182, 0.335]$	$[0.054, 0.145]$
$\widehat{t}(\widehat{\underline{k}}_{LIML})$	$[0.059, 0.145]$	$[0.044, 0.125]$	$(-\infty, \infty)$	$[0.054, 0.145]$
$t_0(\widehat{\underline{k}}_{2SLS})$	$[0.059, 0.145]$	$[0.044, 0.128]$	$[-0.517, 0.919]$	$[0.056, 0.155]$
$\widehat{t}(\widehat{\underline{k}}_{2SLS})$	$[0.059, 0.145]$	$[0.044, 0.128]$	$(-\infty, \infty)$	$[0.056, 0.155]$
$t_0(\widehat{\underline{k}}_{FULL})$	$[0.059, 0.145]$	$[0.044, 0.125]$	$[-0.197, 0.334]$	$[0.054, 0.145]$
$\widehat{t}(\widehat{\underline{k}}_{FULL})$	$[0.059, 0.145]$	$[0.044, 0.125]$	$(-\infty, \infty)$	$[0.054, 0.145]$
$t_0(\widehat{\underline{k}}_{B2SLS})$	$[0.059, 0.145]$	$[0.044, 0.128]$	$(-\infty, \infty)$	$[0.055, 0.136]$
$\widehat{t}(\widehat{\underline{k}}_{B2SLS})$	$[0.059, 0.145]$	$[0.048, 0.123]$	$[-0.384, 0.505]$	$[0.070, 0.155]$
F(first stage)	30.528	4.748	1.613	1.870
k	3	30	28	178

Table 3: Effects of Years of Education on Log Weekly Earnings (1940-49 cohort)

	I	II	III	IV
$\widehat{\beta}(\widehat{\underline{k}}_{OLS})$ $\underset{(s.e)}{}$	0.0632 (0.003)	0.0632 (0.003)	0.0632 (0.003)	0.0628 (0.003)
$\widehat{\beta}(\widehat{\underline{k}}_{LIML})$ $\underset{(s.e)}{}$	-0.0902 (0.0301)	0.0286 (0.0197)	0.1243 (0.0420)	0.0878 (0.0178)
$\widehat{\beta}(\widehat{\underline{k}}_{2SLS})$ $\underset{(s.e)}{}$	-0.0734 (0.0273)	0.0393 (0.0145)	0.0779 (0.0239)	0.0666 (0.0113)
$\widehat{\beta}(\widehat{\underline{k}}_{FULL})$ $\underset{(s.e)}{}$	-0.0882 (0.0299)	0.0289 (0.0196)	0.1218 (0.0412)	0.0875 (0.0177)
$\widehat{\beta}(\widehat{\underline{k}}_{B2SLS})$ $\underset{(s.e)}{}$	-0.0750 (0.0275)	0.0373 (0.0156)	0.0912 (0.0297)	0.0872 (0.0164)
	1-sided 95% Confidence Regions			
$CLR1$	$[-0.148, \infty)$	$[-0.008, \infty)$	$[0.043, \infty)$	$[0.046, \infty)$
$CMLR1$	$[-0.148, \infty)$	$[-0.008, \infty)$	$[0.043, \infty)$	$[0.046, \infty)$
$t_0(\widehat{\underline{k}}_{LIML})$	$[-0.148, \infty)$	$[-0.008, \infty)$	$[0.043, \infty)$	$[0.046, \infty)$
$\widehat{t}(\widehat{\underline{k}}_{LIML})$	$[-0.148, \infty)$	$[-0.007, \infty)$	$[0.041, \infty)$	$[0.046, \infty)$
$t_0(\widehat{\underline{k}}_{2SLS})$	$[-0.130, \infty)$	$[0.009, \infty)$	$[0.031, \infty)$	$[0.044, \infty)$
$\widehat{t}(\widehat{\underline{k}}_{2SLS})$	$[-0.130, \infty)$	$[0.009, \infty)$	$[0.031, \infty)$	$[0.044, \infty)$
$t_0(\widehat{\underline{k}}_{FULL})$	$[-0.148, \infty)$	$[-0.008, \infty)$	$[0.043, \infty)$	$[0.046, \infty)$
$\widehat{t}(\widehat{\underline{k}}_{FULL})$	$[-0.148, \infty)$	$[-0.007, \infty)$	$[0.040, \infty)$	$[0.046, \infty)$
$t_0(\widehat{\underline{k}}_{B2SLS})$	$[-0.130, \infty)$	$[0.008, \infty)$	$[0.031, \infty)$	$[0.044, \infty)$
$\widehat{t}(\widehat{\underline{k}}_{B2SLS})$	$[-0.130, \infty)$	$[0.011, \infty)$	$[0.043, \infty)$	$[0.055, \infty)$
	2-sided 95% Confidence Regions			
CLR	$[-0.161, -0.036]$	$[-0.015, 0.071]$	$[0.026, 0.262]$	$[0.038, 0.142]$
$t_0(\widehat{\underline{k}}_{LIML})$	$[-0.161, -0.036]$	$[-0.015, 0.070]$	$[0.028, 0.264]$	$[0.037, 0.141]$
$\widehat{t}(\widehat{\underline{k}}_{LIML})$	$[-0.160, -0.036]$	$[-0.015, 0.070]$	$[0.024, 0.256]$	$[0.037, 0.141]$
$t_0(\widehat{\underline{k}}_{2SLS})$	$[-0.141, -0.026]$	$[0.003, 0.070]$	$[0.019, 0.179]$	$[0.036, 0.133]$
$\widehat{t}(\widehat{\underline{k}}_{2SLS})$	$[-0.141, -0.026]$	$[0.003, 0.070]$	$[0.019, 0.178]$	$[0.036, 0.133]$
$t_0(\widehat{\underline{k}}_{FULL})$	$[-0.161, -0.036]$	$[-0.015, 0.070]$	$[0.027, 0.263]$	$[0.037, 0.141]$
$\widehat{t}(\widehat{\underline{k}}_{FULL})$	$[-0.160, -0.036]$	$[-0.015, 0.070]$	$[0.023, 0.255]$	$[0.037, 0.141]$
$t_0(\widehat{\underline{k}}_{B2SLS})$	$[-0.142, -0.025]$	$[0.003, 0.071]$	$[0.019, 0.180]$	$[0.036, 0.134]$
$\widehat{t}(\widehat{\underline{k}}_{B2SLS})$	$[-0.141, -0.026]$	$[0.005, 0.068]$	$[0.034, 0.160]$	$[0.050, 0.117]$
F(first stage)	26.316	6.849	2.736	1.929
k	3	30	28	178

Chapter 2

Maxmin and Minimax Regret tests satisfying general constraints

2.1 Introduction

The classic theory of hypothesis testing defines the *best* optimal test as the test that uniformly maximizes the power function among all α-level tests. Such test is called a uniformly most powerful (UMP) test. The existence of a UMP test, however, is restricted to specific hypothesis testing problems. For example, in a Gaussian linear regression model with known variance, there is a UMP test for the one-sided hypothesis about linear combinations of the slope. In the two-sided hypothesis, however, there is no UMP test (see Lehmann and Romano (2005) for a comprehensive introduction and references).

The most common procedures for finding optimal tests when the testing problem does not have a UMP are: 1) consider a different set of tests than the set of all α-level tests; 2) define a prior distribution for the alternative set, and maximize the weighted average power (WAP); 3) use either a maxmin or a minimax regret criterion. The first procedure considers a different set of tests than the set of all α-level tests, but with desirable properties such as unbiasedness, local unbiasedness, invariance, similarity, and others. As in the UMP case, there are specific problems inherent to a UMP test for the new restricted class of tests. For example, the two-sided hypotheses for the linear combination of the slope in the Gaussian linear model have a UMP test over the set of unbiased tests. For more general hypothesis testing problems, however, the existence of UMP tests over the new set of tests is not guaranteed.

The second and the third method relax the restriction of maximizing the power function uniformly over the set of alternatives, and then ensure the existence, for each method, of optimal tests in general problems. In the second approach, the optimal test maximizes the WAP with respect to some set of weights under the alternative hypothesis. In this approach the optimal test is known as a WAP test. A drawback of this procedure is that the applied researcher has to choose the prior distribution for the alternative hypothesis.

The last approach considers two optimality criteria, the maxmin and the minimax regret, and does not require a choice of a prior distribution for the alternative hypothesis. In the maxmin criterion, the optimal test maximizes the lowest power under the alternative hypothesis. In the minimax regret criterion, the optimal test minimizes the largest distance between the power envelope and the power function under the alternative hypothesis. We call these optimal test as a maxmin test and minimax regret test, respectively.

Recently, Moreira and Moreira (2013) (hereinafter, MM) consider optimal tests that combine both the first and the second methods. The use of the second method guarantees the existence of the WAP test, and the first method ensures a WAP test with desirable properties. Specifically, they propose to use tests maximizing the WAP over a general set of tests, that accommodates restrictions like similarity, unbiasedness, local unbiasedness and others.

In this chapter we consider the maxmin and minimax regret criteria of optimality and, as in MM, we use the general set of tests to obtain tests with desirable properties. We call these tests as general maxmin (GM) and general minimax regret (GMR) tests, respectively. Maxmin tests are widely used in hypothesis testing when there is no UMP test (see, for example, Wald (1945), Lehamnn (1952), Krafft and Witting (1967),

Witting (1985), Vajda (1989), Cvitanic and Karatzas (2001), Lehmann and Romano (2005), Rudloff and Karatzas (2010), Gushchin (2015)). In particular testing problems, the analytical form of the maxmin test is known, but for broader hypothesis testing its functional form is unknown. In the general case, Krafft and Witting (1967) (hereinafter, KW) prove that the maxmin test is a Bayes test when an auxiliary problem, the dual problem, has a solution. Furthermore, KW prove the existence of the dual solution when the null and alternative hypotheses are finite. More recent papers of Cvitanic and Karatzas (2001), Rudloff and Karatzas (2010) and Gushchin (2015) give different sets of sufficient conditions for the maxmin test to be a Bayes test. Since their assumptions involve topological and algebraic properties about the set of *densities* instead of the set of *parameters*, their results have no direct application in parametric hypothesis testing.

Our first contribution is to prove the existence of the GM test and give a different set of sufficient conditions for the GM test to be a Bayes test with particular priors. A direct corollary of our result is that the maxmin test is a Bayes test for our regularity conditions. Such sufficient conditions are satisfied when the null and alternative hypotheses are finite and the densities are continuous with respect to the parameters. Hence our result extends KW's result.

An alternative optimality criterion for the maxmin is the minimax regret (or most stringent), proposed by Savage (1951). The minimax regret criterion in hypothesis testing was first used by Schaafsma (1966), to search for minimax regret tests in a Gaussian family where the alternative hypothesis is restricted to inequality constraints. See Schaafsma and Smid (1966), Schaafsma (1951), Tsai and Sen (1993) and references therein for extensions within the Gaussian model. Our second contribution is to prove the existence of the GMR test and give sufficient conditions for the GMR test to be a Bayes test. The following corollary shows that the minimax regret test exists and it is a Bayes test under our regularity conditions.

In the last section of this chapter, we consider testing problems that are invariant to some group of transformations. The Hunt-Stein Theorem allows us to restrict the search of maxmin and minimax regret tests to tests that are invariant. We show that this result still holds under the general constraints considered and, so we can simplify the testing problem under general constraints using the invariance principle.

The present chapter is organized as follows. In section 2 we present the general setup for the hypothesis testing problem. In section 3 we present the GM and GMR tests. Sections 4 and 5 prove the existence of the GM and GMR tests and their correspondingly Bayes representation. The following corollaries specialize the GM and GMR results for the maxmin and minimax regret tests. In section 6 we extend the Hunt-Stein Theorem to maxmin and minimax regret tests satisfying the general constraints. An appendix contains all proofs.

2.2 The Hypothesis Testing Problem

Let $(\Omega, \mathcal{A})$ be a measurable space and $\mathcal{P} := \{\mathbb{P}_\theta : \theta \in \Theta\}$ a family of parametric probability measures on $(\Omega, \mathcal{A})$, where the metric space (Θ, d) is split into two disjoint subsets Θ_H and Θ_K, the null and alternative hypotheses. Suppose the existence of a σ-finite measure μ that dominates $\mathcal{P}$ and define the μ Radon-Nikodym densities $f_\theta := d\mathbb{P}_\theta / d\mu$.

In this chapter we are interested in testing whether the parameter θ either belongs to the null hypothesis $H : \theta \in \Theta_H$ or to the alternative hypothesis $K : \theta \in \Theta_K$. For this we choose a test function $\phi \in \Phi := \{\phi \in L^\infty_\mu : \phi(w) \in [0,1]\}$, where $L^\infty_\mu := L^\infty(\Omega, \mathcal{A}, \mu)$ is the Banach space of all essentially bounded measurable functions, such that $\phi(w)$ is the conditional probability of rejecting the null hypothesis H conditional on observing w.

The maxmin criterion of optimality chooses a test function $\phi^* \in \Phi$ within the set of all α-level tests $\Phi_{\alpha,H} := \left\{\phi \in \Phi : \int_\Omega \phi f_\theta d\mu \le \alpha \text{ for all } \theta \in \Theta_H, \alpha \in (0,1)\right\}$, that attains the higher lowest power in the alternatives. Then the maxmin test $\phi^* \in \Phi_{\alpha,H}$ is such that:

$$\inf_{\theta \in \Theta_K} \int_\Omega \phi^* f_\theta d\mu = \sup_{\phi \in \Phi_{\alpha,H}} \inf_{\theta \in \Theta_K} \int_\Omega \phi f_\theta d\mu. \tag{2.1}$$

The seminal paper of KW shows that the maxmin test ϕ^* is a Bayes test (with a Neyman-Pearson structure):

$$\phi^* = \begin{cases} 1 & \text{if } \int_{\Theta_K} f_\theta d\lambda^* > \int_{\Theta_H} f_\theta d\eta^* \\ 0 & \text{if } \int_{\Theta_K} f_\theta d\lambda^* < \int_{\Theta_H} f_\theta d\eta^* \end{cases}, \tag{2.2}$$

if the prior measures λ^* and η^* are solutions of an auxiliary problem, the dual problem. Under the assumption that the parameter sets Θ_K and Θ_H are finite, KW proves the existence of the dual problem. More recently, Cvitanic and Karatzas (2001), Rudloff and Karatzas (2010) and Gushchin (2015) consider the nonparametric testing problem:

$$H' : f \in \mathcal{F}_H := \{\ f_\theta : \theta \in \Theta_H\} \text{ and } K' : f \in \mathcal{F}_K := \{\ f_\theta : \theta \in \Theta_K\}. \tag{2.3}$$

Rudloff and Karatzas (2010) generalize the result of Cvitanic and Karatzas (2001) and prove that: if *RK1)* the set of *densities* in the alternative hypothesis, $\mathcal{F}_K$, is weakly compact and convex; and *RK2)* the set of *densities* in the null hypothesis, $\mathcal{F}_H$, is weakly compact; then the maxmin test has a Bayes representation:

$$\phi^* = \begin{cases} 1 \text{ if } \int_{\mathcal{F}_K} f_\theta d\kappa > \int_{\mathcal{F}_H} f_\theta d\vartheta \\ 0 \text{ if } \int_{\mathcal{F}_K} f_\theta d\kappa < \int_{\mathcal{F}_H} f_\theta d\vartheta \end{cases}, \tag{2.4}$$

where ϑ and κ are priors over $\mathcal{F}_H$ and $\mathcal{F}_K$. Gushchin (2015) extends this result, proving it under a weaker set of sufficient conditions: *G1)* $\{\min\{f_\theta, f_{\theta'}\} : f_\theta \in co(\mathcal{F}_H),\ f_{\theta'} \in co(\mathcal{F}_K)\}$ is uniformly integrable with respect to μ, where $co(A)$ is the convex hull of the set A. Condition *G1)* holds if either one of the families $\mathcal{F}_H$ or $\mathcal{F}_K$ is uniformly integrable. Both results, however, have limited applicability in parametric hypothesis testing, since there is no clear topological and algebraic connection between the set of densities $(\mathcal{F}_H, \mathcal{F}_K)$ and the set of parameters (Θ_H, Θ_K). Consider, for example, the hypothesis testing:

$$H : \zeta = \zeta_0,\ \sigma \in [\sigma_1, \sigma_2] \text{ against } K : \zeta \in [\zeta_1, \zeta_2],\ \sigma \in [\sigma_1, \sigma_2],$$

where ζ and σ are the mean and variance in the normal family. Note that the null parameter set $\Theta_H = \{(\zeta, \sigma) \in \mathbb{R} \times \mathbb{R}_+, \zeta = \zeta_0, \sigma \in [\sigma_1, \sigma_2]\}$ and the alternative parameter set $\Theta_K = \{(\zeta, \sigma) \in \mathbb{R} \times \mathbb{R}_+, \zeta \in [\zeta_1, \zeta_2], \sigma \in [\sigma_1, \sigma_2]\}$ are convex sets. However, the alternative density set:

$$\mathcal{F}_K = \left\{ f_{\zeta,\sigma}(x) := \frac{1}{\sigma} f\left(\frac{x - \zeta}{\sigma}\right) : f \text{ standard normal density, } \zeta \in [\zeta_1, \zeta_2], \sigma \in [\sigma_1, \sigma_2] \right\},$$

is not convex. Indeed, a convex combination of elements in $\mathcal{F}_K$ gives a mixing density, which is not normal and does not belong to $\mathcal{F}_K$. The uniformly integrability condition *G1)* does not hold either. Indeed, for any $\varepsilon \in (0, 1)$ and any $K > 0$ we can choose $\sigma > 0$ such that:

$$\int_{\mathbb{R}} f_{\zeta,\sigma}(x) 1_{\{f_{\zeta,\sigma}(x) > K\}} dx = \int_{\mathbb{R}} 1_{\{z^2 \le c(K,\sigma)\}} f_{0,1}(z) dz > \varepsilon,$$

where Z standard normal variable and $c(K, \sigma) := -2\ln((2pi)^{1/2} K\sigma)$, because the right hand side is the cumulative distribution function of a *chi-squared* distribution with one degree of freedom.

The second optimality criterion considered in this chapter and in the third chapter is the minimax regret. In the minimax regret criterion, the optimal test minimizes the largest distance between the power envelope and the power function. Hence, the minimax regret test $\phi^\dagger \in \Phi_{\alpha,H}$ is such that:

$$\sup_{\theta \in \Theta_K} \left(\beta(\theta) - \int_\Omega \phi^\dagger f_\theta d\mu \right) = \inf_{\phi \in \Phi_{\alpha,H}} \sup_{\theta \in \Theta_K} \left(\beta(\theta) - \int_\Omega \phi f_\theta d\mu \right), \tag{2.5}$$

where $\beta(\theta) := \left(\sup_{\phi \in \Phi_{\alpha,H}} \int_\Omega \phi f_\theta d\mu \right)$ is the power envelope. The previous literature using the minimax regret approach in testing problems is limited, to our knowledge, to the Gaussian family, see for example Schaafsma (1966), Schaafsma and Smid (1966), Schaafsma (1968), Tsai and Sen (1993). Schaafsma and Smid (1966) and Schaafsma (1968) derive minimax regret tests under α-level and broader constraints for the testing problem:

$$H : \zeta = 0 \text{ against } K : \zeta \gneq 0,$$

where $\zeta \in \mathbb{R}^k$ is the mean vector of a Gaussian distribution with known covariance matrix. Tsai and Sen (1993) prove that the minimax regret test (among the set of tests with α-level and with at least one alternative

hypothesis such that the power envelope and the power function are equal) has a better performance than the likelihood ratio test in terms of local power. This result highlights the importance of combining the minimax regret (and the maxmin) optimality criteria with more general constraints to obtain good tests.

Finally, characterizing optimal tests in the finite sample model $\mathcal{F} := \{f_\theta : \theta \in \Theta\}$ is also an important step for finding asymptotically optimal tests. The LeCam asymptotic theory breaks down the search for asymptotically optimal tests into three steps: the first step establishes the convergence of the sequence of models to a limit model; the second step finds the optimal test in the limit model; the last step gives tests with power functions that converge to the power function of the optimal tests in the limit model (see Strasser (1985) and LeCam (1986))[1]. The second step in the search for asymptotically optimal tests becomes the finite sample problem considered in this chapter, where $\mathcal{F}$ is the limit model obtained in the first step.

2.3 The General Maxmin and Minimax Regret problems

In this section, we generalize the usual maxmin and minimax regret problems. We consider two generalizations. First, the tests need to satisfy more general constraints than the α-level restriction, such as unbiasedness, local unbiasedness and similarity. Second, the power functions in the maxmin and the minimax approaches are controlled over a subset of the alternative hypothesis. Both methods give sufficient flexibility to obtain tests with desirable properties and to eliminate trivial known tests. Consider, for example, a hypothesis testing problem where the null and alternative hypotheses have a common boundary $\overline{\Theta}_H \cap \overline{\Theta}_K \neq \emptyset$, where $\overline{A}$ is the closure of the set A. If the power function of any test is continuous, then the trivial test $\phi_0 = \alpha$ is a maxmin test. First, the maxmin test ϕ^* exists by previous literature (see also corollary 13 below). Now note that a maxmin test ϕ^* is unbiased. Indeed,

$$\inf_{\theta \in \Theta_K} \int_\Omega \phi^* f_\theta d\mu \geq \alpha,$$

because $\phi_0 = \alpha \in \Phi_{\alpha,H}$. Note also that the power at the boundary $\theta \in \overline{\Theta}_H \cap \overline{\Theta}_K$ is equal to α, since we can take sequences $\{\theta_n\} \subseteq \Theta_H$ and $\{\theta'_n\} \subseteq \Theta_K$ such

$$\lim_{d(\theta_n,\theta)\to 0} \int_\Omega \phi^* f_{\theta_n} d\mu \ \leq \alpha \leq \lim_{d(\theta'_n,\theta)\to 0} \int_\Omega \phi^* f_{\theta'_n} d\mu.$$

Then, the lowest power is bounded above by α, because

$$\inf_{\theta \in \Theta_K} \int_\Omega \phi^* f_\theta d\mu \leq \int_\Omega \phi^* f_{\theta'_n} d\mu \to \alpha,$$

as $d(\theta'_n, \theta) \to 0$. Finally, the trivial test $\phi_0 = \alpha$ has lowest power equal to α, and as a consequence, it is a maxmin test.

Now we present the generalizations of the problems. The first, proposed by MM, imposes that the tests ϕ have to satisfy the general constraints:

$$\gamma_{k,1}(\theta) \leq \int_\Omega \phi g_\theta^k d\mu \leq \gamma_{k,2}(\theta) \text{ for all } \theta \in \Theta_{k,S},\ k = 1, ..., l, \tag{2.6}$$

where $(\Theta_{k,S})_{k=1}^l$ disjoint subsets of Θ, $g = (g^1, ..., g^l) : \times_{k=1}^l (\Theta_{k,S} \times \Omega) \to \mathbb{R}^l$ measurable function such $g_\theta^k(\cdot) \in L_\mu^1$ for all $\theta \in \Theta_{k,S}$ and $k = 1, ..., l$, where $L_\mu^1 := L^1(\Omega, \mathcal{A}, \mu)$ is the Banach space of all integrable functions; and $\gamma := \{(\gamma_{k,1}, \gamma_{k,2})_{k=1}^l : \times_{k=1}^l (\Theta_{k,S})^2 \to \mathbb{R}^{2l} : \gamma_{k,1}(\theta) \leq \gamma_{k,2}(\theta) \text{ for all } \theta \in \Theta_{k,S} \text{ and } k = 1, ..., l\}$ restriction functions. The general restrictions (2.6) can accommodate the α-level constraint:

$$\int_\Omega \phi f_\theta d\mu \leq \alpha \text{ for all } \theta \in \Theta_H,$$

the unbiasedness constraint:

$$\sup_{\theta \in \Theta_H} \int_\Omega \phi f_\theta d\mu \leq \alpha \leq \inf_{\theta \in \Theta_K} \int_\Omega \phi f_\theta d\mu,$$

[1] The literature also uses the terms sequence of experiments and limit experiment for sequence of models and limit model, respectively.

the similar $\pm\delta$ constraint:

$$\alpha - \delta \leq \int_\Omega \phi f_\theta d\mu \leq \alpha + \delta \text{ for all } \theta \in \Theta_S,$$

for $\Theta_S \subset \Theta$.

The second generalization imposes that the power functions in the maxmin and the minimax regret criteria are controlled over a subset Θ_R of the alternative hypothesis Θ_K. Then, the general maxmin problem (GMP) and the general minimax regret problem (GMRP) are:

$$\sup_{\phi \in \Gamma(g,\gamma)} \inf_{\theta \in \Theta_R} \int_\Omega \phi f_\theta d\mu, \text{ and} \tag{2.7}$$

$$\inf_{\phi \in \Gamma(g,\gamma)} \sup_{\theta \in \Theta_R} \left(\beta^*(\theta) - \int_\Omega \phi f_\theta d\mu \right), \tag{2.8}$$

where $\Gamma(g,\gamma) := \cap_{k=1}^{l} \left\{ \phi \in \Phi : \gamma_{k,1}(\theta) \leq \int_\Omega \phi g_\theta^k d\mu \leq \gamma_{k,2}(\theta) \text{ for all } \theta \in \Theta_{k,S} \right\}$ is the general set of tests and $\beta^*(\theta) := \left(\sup_{\phi \in \Gamma(g,\gamma)} \int_\Omega \phi f_\theta d\mu \right)$ is the general power envelope. In the rest of the chapter, we refer to Θ_R as the alternative set.

Usually the maxmin and minimax regret tests are different. When the power envelope is constant, however, a maxmin test is trivially a minimax regret test. In the following proposition we extend this idea.

Proposition 11 *Suppose that exist a test $\phi^* \in \Gamma(g,\gamma)$ and a partition $\{\Theta_e\}_{e \in E}$ of Θ_R such for each $e \in E$: i) the power envelope is constant on Θ_e, $\beta^*(\theta) = \beta_e$; and ii) ϕ^* is a GM for all Θ_e:*

$$\inf_{\theta \in \Theta_e} \int_\Omega \phi^* f_\theta d\mu = \sup_{\phi \in \Gamma(g,\gamma)} \inf_{\theta \in \Theta_e} \int_\Omega \phi f_\theta d\mu.$$

Then ϕ^* is a GMR test.

The previous proposition shows that when exist a partition of the alternative set Θ_R such that the power envelope is constant on each element of the partition and a GM test ϕ^* is a GM test for each element of the partition, then the GM test ϕ^* is also a GMR test.

2.4 Characterization of the General Maxmin Test

In this section we present a new set of sufficient conditions for the characterization of the GM test as a Bayes test. Furthermore, we show that the priors come from a generalization of the KW's dual problem. The set of sufficient conditions are:

Assumption 1: There exist a test $\overline{\phi} \in \Gamma(g,\gamma)$ such $\overline{\phi}(w) \in (0,1)$ for all $w \in \Omega$, and $\gamma_{1,k}(\theta) < \int_\Omega \overline{\phi} g_\theta^k d\mu < \gamma_{2,k}(\theta)$ for all $\theta \in \Theta_{k,S}$ and $k = 1, ..., l$;

Assumption 2: Θ_R and $\Theta_{k,S}$, $k = 1, ..., l$, are compact sets;

Assumption 3: For every $h \in \{f_\theta \in \mathcal{F}_K : \theta \in \Theta_R\} \cup \{g^k\}_{k=1}^{l}$, the function $\theta \mapsto h(\theta, w)$ is continuous for all $w \in \Omega$, and the functions $\gamma_{k,1}(\cdot)$ and $\gamma_{k,2}(\cdot)$ are continuous for each $k = 1, ..., l$; and

Assumption 4: The functions $\theta \mapsto \int_\Omega g_\theta^k d\mu$ are continuous for all $k = 1, ..., l$.

The first assumption, known as a generalized Slater condition, imposes the existence of a test with non-binding restrictions. For the α-level and similar $\pm\delta$ constraints, for example, the tests $\widetilde{\phi}_0 = \alpha/2$ and $\phi_0 = \alpha$ satisfy these conditions, respectively. The second assumption is more restrictive since it rules out important hypothesis testing problems, for example the two-sided hypotheses $H : \theta = \theta_0$ and $K : \theta \neq \theta_0$, $\theta \in \mathbb{R}$. We

can, however, restrict the parameter set in order to guarantee compactness. In the two-sided hypothesis, for example, we expect that desirable tests have good performance far from the null hypothesis $H : \theta = \theta_0$, hence we could set $\Theta_R = [\theta_1, \theta_0 - \varepsilon] \cup [\theta_0 + \varepsilon, \theta_2]$, for $\varepsilon > 0$ small and $|\theta_1|$ and θ_2 large. Assumptions 3 and 4 are continuity assumptions about the densities and the constraint functions. Maxmin and minimax regret problems using continuously density functions trivially satisfy assumptions 3 and 4.

The proof of the following theorem for the GM test contains two steps: first, we follow closely Krafft and Witting (1967) in stating the GMP as an infinite linear problem and find it's Fenchel-Rockafellar dual problem. In the second step, we use convexity optimization to show that the dual problem has a solution (strong duality property) and the equality between the optimal value of the dual and primal problems (no gap property). For references to duality methods and convex optimization, see Rockafellar (1974), Anderson and Nash (1987), Ekeland and Teman (1999) and the references therein. Let $\mathcal{M}(\mathcal{X})$ denote the space of all regular Borel measures on the Borel σ-algebra of some metric space $\mathcal{X}$; $\mathcal{M}_1(\mathcal{X}) := \{\lambda \in \mathcal{M}(\mathcal{X}) : \lambda(\mathcal{X}) = 1\}$ and $(x)^+ := \max\{x, 0\}$.

Theorem 12 *a) Suppose that $\Gamma(g, \gamma) \neq \emptyset$, then a GM test ϕ^* exists.*

b) Suppose that Assumptions (1)-(4) *hold, then ϕ^* is a GM test if and only if*

$$\phi^* = \begin{cases} 1 & \text{if } \int_{\Theta_R} f_\theta d\lambda^* > \sum_{k=1}^{l} \int_{\Theta_{k,S}} g_\theta^k d(\eta_{k,2}^* - \eta_{k,1}^*) \\ 0 & \text{if } \int_{\Theta_R} f_\theta d\lambda^* < \sum_{k=1}^{l} \int_{\Theta_{k,S}} g_\theta^k d(\eta_{k,2}^* - \eta_{k,1}^*) \end{cases}, \tag{2.9}$$

such that for all $k = 1, ..., l$,

$$\inf_{\theta \in \Theta_R} \int_\Omega \phi^* f_\theta d\mu = \int_\Omega \phi^* f_{\theta'} d\mu \quad \lambda^*\text{-a.e. } \theta' \in \Theta_R \tag{2.10}$$

$$\int_\Omega \phi^* g_\theta^k d\mu = \gamma_{1,k}(\theta') \quad \eta_{k,1}^*\text{-a.e. } \theta' \in \Theta_{k,S} \tag{2.11}$$

$$\int_\Omega \phi^* g_\theta^k d\mu = \gamma_{2,k}(\theta') \quad \eta_{k,2}^*\text{-a.e. } \theta' \in \Theta_{k,S} \tag{2.12}$$

where $\left(\lambda^*, (\eta_{k,1}^*, \eta_{k,2}^*)_{k=1}^l\right) \in (\Lambda, \Xi) := (\mathcal{M}_1(\Theta_R), \times_{k=1}^l \mathcal{M}(\Theta_{k,S})^2)$ is a solution of

$$\inf_{(\lambda, (\eta_{k,1}, \eta_{k,2})_{k=1}^l) \in (\Lambda, \Xi)} \sum_{k=1}^{l} \left(\int_{\Theta_{k,S}} \gamma_{k,2}(\theta) d\eta_{k,2} - \int_{\Theta_{k,S}} \gamma_{k,1}(\theta) d\eta_{k,1} \right) + \int_\Omega \left(\int_{\Theta_R} f_\theta d\lambda - \sum_{k=1}^{l} \int_{\Theta_{k,S}} g_\theta^k d(\eta_{k,2} - \eta_{k,1}) \right)^+ d\mu \tag{2.13}$$

Theorem 12 a) establishes the existence of the GM test ϕ^* when the general set of tests is nonempty. Under *Assumptions (1)-(4)*, Theorem 12 b) proves that ϕ^* is a GM test if and only if it has the Bayes representation (2.2) where the probability measure λ^* and the signed measures $\eta_k^* := \eta_{k,2}^* - \eta_{k,1}^*$, $k = 1, ..., l$, come from the dual problem (2.13), a generalization of KW's dual problem. Hence, the GM problem is equivalent to the dual problem (2.13). In chapter 3, we described how to numerically implement the GM test.

A direct consequence of Theorem 12 is the solution of the original maxmin problem.

Corollary 13 *a) A maxmin test ϕ^* always exists.*

b) Suppose that Assumption (2)′ Θ_H and Θ_K are compact; and *Assumption* (3)’ For every $f \in \mathcal{F}$, $\theta \mapsto f_\theta(w)$ is a continuous function for all $w \in \Omega$. *Then ϕ^* is a maxmin test if and only if*

$$\phi^* = \begin{cases} 1 & \text{if } \int_{\Theta_K} f_\theta d\lambda^* > \int_{\Theta_H} f_\theta d\eta^* \\ 0 & \text{if } \int_{\Theta_K} f_\theta d\lambda^* < \int_{\Theta_H} f_\theta d\eta^* \end{cases}, \tag{2.14}$$

such that

$$\inf_{\theta \in \Theta_K} \int_\Omega \phi^* f_\theta d\mu = \int_\Omega \phi^* f_{\theta'} d\mu \quad \lambda^*\text{-a.e. } \theta' \in \Theta_K \tag{2.15}$$

$$\int_\Omega \phi^* f_\theta d\mu = \alpha \quad \eta^*\text{-a.e. } \theta \in \Theta_H \tag{2.16}$$

where $(\lambda^*, \eta^*) \in (\mathcal{M}_1(\Theta_K), \mathcal{M}(\Theta_H))$ *is solution of*

$$\inf_{(\lambda,\eta)\in(\mathcal{M}_1(\Theta_K),\mathcal{M}(\Theta_H))} \alpha\eta(\Theta_H) + \int_\Omega \left(\int_{\Theta_K} f_\theta d\lambda - \int_{\Theta_H} f_\theta d\eta \right)^+ d\mu \tag{2.17}$$

The existence of the Maxmin test, Corollary 13 a), is a well known result in the literature and we present it for sake of clarity. According to the Corollary 13 b), the Maxmin test ϕ^* is a Bayes test where the prior probability for the alternative hypothesis, λ^*, and the prior measure for the null hypothesis, η^*, are solutions of the Dual Problem (2.17).

The maxmin problem (2.1) with a simple alternative hypothesis $K : f = \overline{f}$ is the Neyman-Pearson problem for the most powerful test. To obtain the most powerful test, it is sufficient to find a probability η over Θ_H such that Neyman-Pearson test for the hypothesis testing:

$$H^\eta : f = \int_{\Theta_H} f_\theta d\eta \text{ against } K : f = \overline{f},$$

has correct size for the original null hypothesis $H : \theta \in \Theta_H$. (see Theorem 3.8.1 in Lehmann and Romano (2005)). Such probability, when it exists, is called the least favorable distribution. Under the assumptions that Θ_H is a compact set and the density functions are continuous with respect to the parameter, Corollary 13 proves the existence of the least favorable distribution and its form: $\eta = \eta^*/\eta^*(\Theta_H)$, where $\eta^* > 0$ is the solution of the dual problem (2.17) with $\displaystyle\int_{\Theta_K} f_\theta d\lambda = \overline{f}$. For other sets of sufficient conditions, see for example Lehamnn (1952) and Reinhardt (1961).

2.5 Characterization of the General Minimax Regret Test

Now we present analogous results to Theorem 12 and Corollary 13 of the GM and maxmin tests for the GMR and minimax regret tests. The proof that GMR test is a Bayes test contains two steps: first, we use Sion Minimax Theorem (Sion (1958)) in the GMR problem; second, we use Theorem 13 b) and its no gap property to obtain the dual problem for the GMR problem.

Theorem 14 *a) Suppose that* $\Gamma(g, \gamma) \neq \emptyset$, *then a GMR test* $\phi^\dagger$ *exists.*

b) Suppose that Assumptions (1)-(4) *holds, then* $\phi^\dagger$ *is a GMR test if and only if*

$$\phi^\dagger = \begin{cases} 1 & \text{if } \int_{\Theta_R} f_\theta d\lambda^\dagger > \sum_{k=1}^l \int_{\Theta_{k,S}} g_\theta^k d(\eta_{k,2}^\dagger - \eta_{k,1}^\dagger) \\ 0 & \text{if } \int_{\Theta_R} f_\theta d\lambda^\dagger < \sum_{k=1}^l \int_{\Theta_{k,S}} g_\theta^k d(\eta_{k,2}^\dagger - \eta_{k,1}^\dagger) \end{cases}, \tag{2.18}$$

such that for all $k = 1, ..., l$,

$$\sup_{\theta\in\Theta_R} \left(\beta^*(\theta) - \int_\Omega \phi^\dagger f_\theta d\mu \right) = \beta^*(\theta') - \int_\Omega \phi^\dagger f_{\theta'} d\mu \quad \lambda^\dagger\text{-a.e. } \theta' \in \Theta_R \tag{2.19}$$

$$\int_\Omega \phi^\dagger g_\theta^k d\mu = \gamma_{k,1}(\theta') \quad \eta_{k,1}^\dagger\text{-a.e. } \theta' \in \Theta_{k,S} \tag{2.20}$$

$$\int_\Omega \phi^\dagger g_\theta^k d\mu = \gamma_{k,2}(\theta') \quad \eta_{k,2}^\dagger\text{-a.e. } \theta' \in \Theta_{k,S} \tag{2.21}$$

where $\left(\lambda^{\dagger}, \left(\eta_{k,1}^{\dagger}, \eta_{k,2}^{\dagger}\right)_{k=1}^{l}\right) \in (\Lambda, \Xi)$ *is solution of*

$$\sup_{\left(\lambda, \left(\eta_{k,1}, \eta_{k,2}\right)_{k=1}^{l}\right) \in (\Lambda, \Xi)} \int_{\Theta_R} \beta^*(\theta) d\lambda + \sum_{k=1}^{l} \left(\int_{\Theta_{k,S}} \gamma_{k,1}(\theta) d\eta_{k,1} - \int_{\Theta_{k,S}} \gamma_{k,2}(\theta) d\eta_{k,2} \right)$$
$$- \int_{\Omega} \left(\int_{\Theta_R} f_\theta d\lambda - \sum_{k=1}^{l} \int_{\Theta_{k,S}} g_\theta^k d(\eta_{k,2} - \eta_{k,1}) \right)^{+} d\mu \qquad (2.22)$$

Theorem 14 (a) proves the existence of the GMR test when the general set of tests is nonempty. Under *Assumptions (1)-(4),* Theorem 14 (b) shows that the test function $\phi^{\dagger}$ is a GMR test if only if it is a Bayes test with prior measures $\lambda^{\dagger}$ and $\eta_k^{\dagger} := \eta_{k,2}^{\dagger} - \eta_{k,1}^{\dagger}$, $k = 1, ..., l$, that are solutions of the dual problem (2.22). Note that both the GM and GMR tests have the same structural form and satisfy the binding constraints 2.20 and 2.21, but the priors λ and $\eta_k = \eta_{k,2} - \eta_{k,1}$, $k = 1, ..., l$, are solutions of different dual problems.

Applying Theorem 12 to the original minimax regret problem, we obtain the following corollary.

Corollary 15 *a) A minimax regret test* $\phi^{\dagger}$ *for the original problem (2.5) always exists.*

b) Suppose that: Assumption (2)' Θ_H and Θ_K are compact; and *Assumption* (3)' For every $f \in \mathcal{F}$, $\theta \mapsto f_\theta(w)$ is a continuous function for all $w \in \Omega$. *Then* $\phi^{\dagger}$ *is a minimax regret test if and only if*

$$\phi^{\dagger} = \begin{cases} 1 & \text{if } \int_{\Theta_K} f_\theta d\lambda^{\dagger} > \int_{\Theta_H} f_\theta d\eta^{\dagger} \\ 0 & \text{if } \int_{\Theta_K} f_\theta d\lambda^{\dagger} < \int_{\Theta_H} f_\theta d\eta^{\dagger} \end{cases}, \qquad (2.23)$$

such that

$$\sup_{\theta \in \Theta_K} \left(\beta(\theta) - \int_{\Omega} \phi^{\dagger} f_\theta d\mu \right) = \beta(\theta') - \int_{\Omega} \phi^{\dagger} f_{\theta'} d\mu \quad \lambda^{\dagger}\text{-a.e. } \theta' \in \Theta_K \qquad (2.24)$$

$$\int_{\Omega} \phi^{\dagger} f_\theta d\mu = \alpha \quad \eta^{\dagger}\text{-a.e. } \theta \in \Theta_H \qquad (2.25)$$

where $(\lambda^{\dagger}, \eta^{\dagger}) \in (\mathcal{M}_1(\Theta_K), \mathcal{M}(\Theta_H))$ *is solution of*

$$\sup_{(\lambda, \eta) \in (\mathcal{M}_1(\Theta_K), \mathcal{M}(\Theta_H))} \int_{\Theta_R} \beta(\theta) d\lambda - \alpha \eta(\Theta_H) - \int_{\Omega} \left(\int_{\Theta_K} f_\theta d\lambda - \int_{\Theta_H} f_\theta d\eta \right)^{+} d\mu \qquad (2.26)$$

Corollary 15 (a) is, to our knowledge, the first existence result for the minimax regret test under general hypothesis testing problems. Corollary 15 (b) presents the Bayes representation of the minimax regret, with priors $\lambda^{\dagger}$ and $\eta^{\dagger}$ that are solutions of the dual problem in corollary (15).

2.6 Invariant Hypothesis Testing

In this section we consider testing problems that are invariant to some group of transformations. Under invariance of the model, the Hunt-Stein Theorem allows us to restrict the search for maxmin and minimax regret tests to invariant tests. We accommodate this result for maxmin and minimax tests satisfying the general constraints. Such result relaxes the compactness and continuity assumptions of the previous sections and can simplify the testing problem by the invariance principle.

Let $(G, \mathfrak{G})$ be a locally compact Hausdorff group and $\{u_\tau : \Omega \to \Omega : \text{measurable transformation}, \tau \in G\}$ such that $u_{\tau_1} \circ u_{\tau_2} = u_{\tau_1 \circ \tau_2}$ for all $\tau_1, \tau_2 \in G$. Under locally compactness, there exists right Haar measure μ_H on $(G, \mathfrak{G})$: μ_H is a Radon measure such $\mu_H(O) > 0$ for all open $O \in \mathfrak{G}$ and $\mu_H(A\tau) = \mu_H(A)$ for all $A \in \mathfrak{G}$ (see Theorem 1, n.2 in Bourbaki (2004)). The probability family $\mathcal{P}$ is invariant with respect to $(G, \mathfrak{G})$ if $\mathbb{P}_\theta \circ u_\tau^{-1} \in \mathcal{P}$ for all $\theta \in \Theta$ and $\tau \in G$. Under identification of the parameter $\theta \in \Theta$, the set of transformations $\{v_\tau : \Theta \to \Theta : \mathbb{P}_\theta \circ u_\tau^{-1} = \mathbb{P}_{v_\tau(\theta)}\}$ defines a group of transformations over the parameter space. We say that the testing problem is invariant to the group $(G, \mathfrak{G})$ if for all $\tau \in G$:

$$\begin{aligned} v_\tau(\Theta_R) &= \Theta_R \\ v_\tau(\Theta_{k,S}) &= \Theta_{k,S}, \text{ for all } k = 1, .., l, \end{aligned}$$

and the test $\phi \in \Phi$ is invariant if $\phi(u_\tau(w)) = \phi(w)$ for all $\tau \in G$ and $w \in \Omega$. In this section, we make the following assumptions:

Assumption I.1: i) Each constraint functions g^k in $g = (g^1, ..., g^l)$ can be decomposed as $g^k_\theta(w) = m_k(w) f_\theta(w)$ where m^k is a $(G, \mathfrak{G})$ invariant function: $m_k(u_\tau(w)) = m_k(w)$ $\forall \tau \in G$ and $w \in \Omega$; ii) The constraints γ are $(G, \mathfrak{G})$ invariant: $\gamma_{k,1}(v_\tau(\theta)) = \gamma_{k,1}(\theta)$ and $\gamma_{k,2}(v_\tau(\theta)) = \gamma_{k,2}(\theta)$ for all $\tau \in G$ and $k = 1, ..., l$.

Assumption I.2: $(G, \mathfrak{G})$ is amenable: exists a sequence of compacts $K_n \subseteq K_{n+1}$ such that $\cup_{n\in\mathbb{N}} K_n = G$ and:

$$\lim_{n\to\infty} \sup_{A\in\mathfrak{G}} \left| \frac{\mu_H(A\tau \cap K_n)}{\mu_H(K_n)} - \frac{\mu_H(A \cap K_n)}{\mu_H(K_n)} \right| = 0.$$

Assumption I.1 - i) imposes that the constraint functions g^k can be decomposed as a product between the density function f_θ and an invariant function m_k. For testing problems with constraints depending on densities only (e.g. level, unbiasedness, similarity constraints), this assumption is trivially satisfied. For more general constraints, however, we need to check this condition. The locally unbiased restriction, for example, imposes the orthogonality between the test and a specific statistic. This assumption is satisfied if the specific statistic is invariant to the group of transformations. Assumption I.1 - ii) requires that the constraint functions γ_k to be constant over the orbits: $Orb(\theta) := \{v_\tau(\theta) : \tau \in G\}$. Assumption I.2 is the main assumption in the Hunt-Stein Theorem. It is satisfied if G is a locally compact and σ-compact Abelian group (Kerstan and Matthes (1969)). In the following corollary we adapt the Hunt-Stein Theorem to our testing problem.

Corollary 16 *Suppose that the testing problem is invariant to the Group $(G, \mathfrak{G})$ and Assumptions I.(1)-(2) hold. Then every test $\phi \in \Gamma(g, \gamma)$, there exist a $\mathcal{P}$-almost invariant test $\phi_I \in \Gamma(g, \gamma)$ such has a better performance in terms of Maxmin and Minimax Regret criteria:*

$$\begin{aligned} \inf_{\theta\in\Theta_R} \int_\Omega \phi f_\theta d\mu &\leq \inf_{\theta\in\Theta_R} \int_\Omega \phi_I f_\theta d\mu \text{ and} \\ \sup_{\theta\in\Theta_R} \left(\beta^*(\theta) - \int_\Omega \phi_I f_\theta d\mu\right) &\leq \sup_{\theta\in\Theta_R} \left(\beta^*(\theta) - \int_\Omega \phi f_\theta d\mu\right). \end{aligned}$$

Under invariance of the testing problem and Assumptions $I.(1)-(2)$, Corollary (16) allows us to restrict the search for the general maxmin and minimax tests to tests that are invariant. This approach is interesting because every invariant test is a function of the maximal invariant statistic (invariance principle) and, in most cases, the distribution of the maximal invariant depends on a low dimensional parameter. Hence, finding an optimal test under the general constraints for the maximal invariant model can be simpler.

Chapter 3

Numerical Implementation of Maxmin and Minimax Regret Tests satisfying general constraints

3.1 Introduction

In the second chapter, Theorems 12 and 14 established the Bayes representation of the maxmin and minimax regret tests satisfying the general constraints (denoted GM and GMR tests). Finding the prior measures of the GM and GMR tests, however, is a hard task since the dual problems (2.13) and (2.22) are infinite dimensional optimization problems. In the present chapter, we implement the discretization method, proposed by Moreira and Moreira (2013) (hereinafter, MM), in order to construct an optimization problem with a finite number of restrictions. This method is interesting because the GM and GMR optimization problems with a finite number of restrictions can be solved numerically. We justify the discretization of the parameter space by showing that the GM and GMR tests with a finite number of restrictions converge, in terms of power in the alternative set, to the GM and GMR tests as the discretization turns finer.

The present chapter is organized as follows. In section 2 we present the discretization scheme and the GM and GM tests satisfying a finite number of restrictions. We also illustrate a case where the discretization method gives optimal tests, i.e., when the GM and GMR tests satisfying a finite number of restrictions are indeed the GM and GMR tests, respectively. In section 3 we present the convergence of the GM and GMR tests satisfying a finite number of constraints, in terms of power, to the GM and GMR tests, respectively. In section 4 we describe the numerical methods for implementing both tests and a sketch of the algorithm. An appendix at the end of the thesis contains proofs of the results.

Example 1 *Moment Inequality*

Recently, the econometric literature has increased the interest in moment inequality models, like, for example, models with treatment effects or microeconomic models with multiple equilibria (see, for example, Imbens and Manski (2004), Rosen (2007), Andrews and Soares (2007) and Chiburis (2009)). Suppose that we know a moment function $m(\cdot)$:

$$\nu_{\varpi^*} = E_{\varpi^*}[m(Z)],$$

where Z random variable and $\nu_{\varpi^*} \in \mathbb{R}^k$. In the moment inequality problem, we are interest in testing the hypothesis:

$$H : \nu_{\varpi^*} \geq 0 \text{ against } K : \nu_{\varpi^*} \ngeq 0. \tag{3.1}$$

Under asymptotic approximations, this testing problem can be set in a much simpler problem, but still complex. Suppose we have Z_i, $i = 1, ..., n$, independent identical distributed random variables. Hence,

$$T_n = \widehat{\Omega}_n^{-1/2}\sqrt{n}(m_n - \nu_{\varpi^*}) \rightsquigarrow N(0, I_k), \tag{3.2}$$

where $m_n = n^{-1}\sum_{i=1}^{n} m(Z_i)$ and $\widehat{\Omega}_n$ is a consistent estimator for the asymptotic variance of $m(Z_i)$. Therefore, $S_n = \widehat{\Omega}_n^{-1/2}\sqrt{n}m_n$ is asymptotic distributed as $N(\widehat{\Omega}_n^{-1/2}\sqrt{n}v_{\varpi^*}, I_k)$ and the testing the problem (3.1) is asymptotic equivalent to testing:

$$H : \theta \geq 0 \text{ against } K : \theta \ngeq 0, \tag{3.3}$$

where θ is the mean of a normal random vector Y, i.e. $Y \sim N(\theta, I_k)$.

Although the testing problem (3.3) is simpler than the nonparametric testing problem in (3.1), it still has some issues. For example, for $k > 1$ there is no uniformly most powerful (UMP) test. Furthermore, any similar test has power equal to size α at some points in the alternative hypothesis (see MM and Andrews (2012)).

Recently, Chiburis (2009) proposes an algorithm to find weighted average power (WAP) tests for the moment inequality problem. WAP tests, however, are Bayes tests with priors being defined by the researcher and not by an optimality criterion. The advantage of the methods proposed in this thesis is that the Maxmin and the Minimax Regret tests are defined by endogenously priors and are based on optimality criteria (see Chapter 2). In the next sections, we present an algorithm to find the GM and GMR tests for any testing problem and then illustrate it with the moment inequality example.

3.2 Discretization Method

In this section we describe the discretization method proposed by MM. Consider the GM and GMR problems presented in the second chapter:

$$\sup_{\phi\in\Gamma(g,\gamma)} \inf_{\theta\in\Theta_R} \int_\Omega \phi f_\theta d\mu, \text{ and} \tag{3.4}$$

$$\inf_{\phi\in\Gamma(g,\gamma)} \sup_{\theta\in\Theta_R} \left(\beta^*(\theta) - \int_\Omega \phi f_\theta d\mu\right), \tag{3.5}$$

where $\Gamma(g,\gamma) := \cap_{k=1}^{l}\{\phi \in \Phi : \gamma_{k,1}(\theta) \leq \int_\Omega \phi g_\theta^k d\mu \leq \gamma_{k,2}(\theta) \text{ for all } \theta \in \Theta_{k,S}\}$ is the general set of tests; $\Theta_R \subseteq \Theta_K$ is the general alternative set and $(\Theta_{k,S})_{k=1}^{l} \subseteq \Theta$ are the general restriction sets; $\beta^*(\theta) := \sup_{\phi\in\Gamma(g,\gamma)} \int_\Omega \phi f_\theta d\mu$ is the general power envelope, presented in the second chapter. To obtain optimization problems (3.4) and (3.5) with a finite number of restrictions, we partition the general sets Θ_R and $(\Theta_{k,S})_{k=1}^{l}$, and choose a finite number of elements of the partitions. For a fixed integer $n \in \mathbb{N}$, pick partitions

$$P_R^n = \{\Pi_{R,i}^n\}_{i=1}^n \text{ and } P_{k,S}^n = \{\Pi_{k,S,i}^n\}_{i=1}^n,$$

of Θ_R and $\Theta_{k,S}$, for all $k = 1, ..., l$, respectively. Given the partitions, we choose representative parameters in each element of the partitions. Define $\{\theta_{R,i}^n\}_{i=1}^n$ and $\{\theta_{k,S,i}^n\}_{i=1}^n$ as a sequence of elements of Θ_R and $\Theta_{k,S}$, such that $\theta_{R,i}^n \in \Pi_{R,i}^n$ and $\theta_{k,S,i}^n \in \Pi_{k,S,i}^n$ for all $k = 1, ..., l$ and $i = 1, ..., n$. Then this finite number of points in the parameter space Θ_R and $\Theta_{k,S}$, $k = 1, ..., l$, can define GM and GMR problems with a finite number of constraints. The GM problem for the finite number of points $\{\theta_{R,i}^n\}_{i=1}^n$ and $\{\theta_{k,S,i}^n\}_{i=1,k=1}^{n,l}$, denoted n-GMP, is defined as:

$$\sup_{\phi\in\Gamma^n(g,\gamma)} \min_{\theta\in\{\theta_{R,i}^n\}_{i=1}^n} \int_\Omega \phi f_\theta d\mu. \tag{3.6}$$

where $\Gamma^n(g,\gamma) := \cap_{k=1}^{l}\{\phi \in \Phi : \gamma_{k,1}(\theta) \leq \int \phi g_\theta^k d\mu \leq \gamma_{k,2}(\theta) \text{ for all } \theta \in \{\theta_{k,S,i}^n\}_{i=1}^n\}$ is the general set of tests satisfying the constraints only in the finite subset $\{\theta_{k,S,i}^n\}_{i=1,k=1}^{n,l}$. We say that a test $\phi^{n,*}$ solving (3.6) is an n-GM test. Analogously, we define the n-GMRP as the GMR problem with constraints only for thefinite number of points $\{\theta_{R,i}^n\}_{i=1}^n$ and $\{\theta_{k,S,i}^n\}_{i=1,k=1}^{n,l}$:

$$\inf_{\phi\in\Gamma^n(g,\gamma)} \max_{\theta\in\{\theta_{R,i}^n\}_{i=1}^n} \left(\beta^*(\theta) - \int_\Omega \phi f_\theta d\mu\right). \tag{3.7}$$

The n-GMR test $\phi^{n,\dagger}$ is a test that solves the n-GMRP (3.7).

In most cases, the discretization scheme will give n-GM and n-GMR tests that are different than the GM and GMR tests, respectively. In the next proposition, however, we present a special case where the optimal tests satisfying the finite number of constraints are indeed the optimal tests for the entire parameter space.

Proposition 17 *a) Let $\phi^{n,*}$ be a n-GM test, if i) $\phi^{n,*} \in \Gamma(g,\gamma)$ and ii)*

$$\inf_{\theta \in \{\theta^n_{R,i}\}^n_{i=1}} \int_\Omega \phi^{n,*} f_\theta d\mu = \inf_{\theta \in \Theta_R} \int_\Omega \phi^{n,*} f_\theta d\mu, \tag{3.8}$$

then $\phi^{n,*}$ is a GM test.

b) Let $\phi^{n,\dagger}$ be a n-GMR test, if i) $\phi^{n,\dagger} \in \Gamma(g,\gamma)$ and ii)

$$\sup_{\theta \in \{\theta^n_{R,i}\}^n_{i=1}} \left(\beta^*(\theta) - \int_\Omega \phi^{n,\dagger} f_\theta d\mu\right) = \sup_{\theta \in \Theta_R} \left(\beta^*(\theta) - \int_\Omega \phi^{n,\dagger} f_\theta d\mu\right), \tag{3.9}$$

then $\phi^{n,\dagger}$ is a GMR test.

3.3 Power Convergence

In this section we present the main result of the chapter. The following theorem describes the behavior of the n-GM test $\phi^{n,*}$ and the n-GMR test $\phi^{n,\dagger}$ as the discretization scheme considers more restrictions in the optimization problems. More specifically, it describes the power convergence of the n-GM test $\phi^{n,*}$ and the n-GMR test $\phi^{n,\dagger}$ as the partitions converge to zero:

$$||P^n_R|| := \sup_{i=1,\ldots,n} \sup_{\theta,\theta' \in \Pi^n_{R,i}} d(\theta,\theta') \to 0, \text{ and}$$
$$||P^n_S|| := \sup_{i=1,\ldots,n;\ k=1,\ldots,l} \sup_{\theta,\theta' \in \Pi^n_{k,S,i}} d(\theta,\theta') \to 0,$$

when $n \to \infty$. In this chapter, we impose Assumptions (1)-(4) of the second chapter and the additional assumption:

Assumption 5: $\int \sup_{\theta \in \Theta_R} f_\theta(w) d\mu < \infty$, $\int \sup_{\theta \in \Theta_{k,S}} |g^k_\theta(w)| d\mu < \infty$ for all $k = 1,\ldots,l$.

Theorem 18 *Suppose that* Assumptions (1)-(5) *holds and let ϕ^* and $\phi^\dagger$ be GM and GMR tests, respectively.Then the n-GM test $\phi^{n,*}$of (3.6) and the n-GMR test $\phi^{n,\dagger}$of (3.7) satisfy*

$$\inf_{\theta \in \{\theta^n_{R,i}\}^n_{i=1}} \int_\Omega \phi^{n,*} f_\theta d\mu \to \int_\Omega \phi^* f_{\theta'} d\mu \quad \lambda^*\text{-a.e. } \theta' \in \Theta_R \text{ and}$$

$$\sup_{\theta \in \{\theta^n_{R,i}\}^n_{i=1}} \left(\beta^*(\theta) - \int_\Omega \phi^{n,\dagger} f_\theta d\mu\right) \to \left(\beta^*(\theta) - \int_\Omega \phi^\dagger f_{\theta'} d\mu\right) \quad \lambda^\dagger\text{-a.e. } \theta' \in \Theta_R,$$

as $||P^n_R|| \to 0$ and $||P^n_S|| \to 0$.

The previous Theorem proves the convergence of the n-GMP and n-GMRP objective functions to the power function of the GM test and the distance between the power envelope and the power function of the GMR test, respectively. Hence, we can numerically implement the tests for a finite number of restrictions as an approximation for the tests satisfying the general constraints. This result and Proposition 17 motivate the use of a finite number of restrictions in the maxmin and minimax regret problems (3.4) and (3.5). As described in the following section, the optimal tests restricted to a finite number of restrictions are implemented numerically.

3.4 Maxmin and Minimax Regret Algorithms

This section describes the importance sampling technique and how it can be used to implement the n-GMP and n-GMRP tests. The first use of importance sampling in hypothesis testing was done by Dantzig (1963). This method has the advantage of reducing the testing problem to a finite linear problem. Linear problems have great numeric appeal given the many efficient algorithms (e.g., simplex, dual, interior point algorithms) available. More recently, MM apply importance sampling techniques to develop efficient algorithms for weighted average power (WAP) tests under general constraints.

Importance sampling is a Monte Carlo integration method. Suppose we want to calculate the integral:

$$\int_{\Omega} \phi(w)p(w)d\mu, \tag{3.10}$$

where ϕ and p are known functions, we can simulate a sample $\{Y_j : j = 1, ..., m_C\}$ from positive density φ, with support Ω, and use the sample estimator:

$$\frac{1}{m_C}\sum_{j=1}^{m_C} \phi(Y_j)\frac{p(Y_j)}{\varphi(Y_j)}. \tag{3.11}$$

By the Strong Law of Large Numbers, we know that the sample estimator (3.11) converges almost surely to the integral (3.10), as the number of simulations m_C diverges. In consequence, the sample estimator is a good approximation for the integral. Note that we can choose any positive density φ, with support Ω, such that the almost sure convergence still holds. Hence, we can choose the density φ efficiently, such as setting φ as the density that attains the minimum variance of the sample estimator (3.11).

3.4.1 Maxmin Algorithm

Now we apply the importance sampling technique to the n-GMP:

$$\sup_{\phi\in\Gamma^n(g,\gamma)} \min_{i=1,\dots,n} \int_{\Omega} \phi f_{\theta^n_{R,i}} d\mu. \tag{3.12}$$

The integrals in the problem (3.12), the integral in the objective function and the integrals in the constraints, are approximated using importance sampling to obtain the following maxmin linear problem:

$$\max_{0\leq x,\ Ax\leq b} \min_{i=1,..n} x'c_i, \tag{3.13}$$

where

$$\begin{aligned}
x &= [\phi(Y_1), ..., \phi(Y_{m_C})]' \in \mathbb{R}^{m_C}_{+},\\
c_i &= [f_{\theta^n_{R,i}}(Y_1)/\left(\varphi(Y_1)m_C\right), ..., f_{\theta^n_{R,i}}(Y_{m_C})/\left(\varphi(Y_{m_C})m_C\right)]' \in \mathbb{R}^{m_C}_{+}\\
A &= [A_1' : -A_1' : I_{m_C}]' \in \mathbb{R}^{(2nl+m_C)\times m_C},\\
A_1 &= [A_1^{1\prime}, ..., A_1^{l\prime}]' \in \mathbb{R}^{nl\times m_C},\ \text{for each } k = 1, ..., l,\\
A_1^k &= \begin{bmatrix} g_{\theta_{k,S,1}}(Y_1)/\left(\varphi(Y_1)m_C\right) & ... & g_{\theta_{k,S,1}}(Y_{m_C})/\left(\varphi(Y_{m_C})m_C\right)\\ \vdots & \ddots & \vdots \\ g_{\theta_{k,S,n}}(Y_1)/\left(\varphi(Y_1)m_C\right) & ... & g_{\theta_{k,S,n}}(Y_{m_C})/\left(\varphi(Y_{m_C})m_C\right)\end{bmatrix} \in \mathbb{R}^{n\times m_C},\\
I_{m_C} &\in \mathbb{R}^{m_C\times m_C} \text{ identity matrix}, b = [\overline{\gamma}_2', -\overline{\gamma}_1', \iota_{m_C}']' \in \mathbb{R}^{(2nl+m_C)},\ \overline{\gamma}_a = [\overline{\gamma}_{1,a}', ..., \overline{\gamma}_{l,a}']' \in \mathbb{R}^{nl},\\
\overline{\gamma}_{k,a} &= [\gamma_{k,a}(\theta_{k,S,1}), ..., \gamma_{k,a}(\theta_{k,S,n})]' \in \mathbb{R}^n \text{ for } k = 1, .., l \text{ and } a = 1, 2,\ \iota_{m_C} \in \mathbb{R}^{m_C} \text{ vector of ones.}
\end{aligned} \tag{3.14}$$

Such minimax linear problem can be solved by standard algorithms, such as the *fminimax* function in Matlab. We gather this optimization procedure in the following algorithm.

General Maxmin Algorithm: Solving the n-GMP:

1 - Let $n \in \mathbb{N}$ be a large integer. For $i = 1, ..., n$, choose:

$$\theta^n_{R,i} \in \Pi^n_{R,i} \text{ and } \theta^n_{k,S,i} \in \Pi^n_{k,S,i} \text{ for all } k = 1, ..., l,$$

where $\{\Pi^n_{R,i}\}^n_{i=1}$ and $\{\Pi^n_{k,S,i}\}^n_{i=1}$ partitions of Θ_R and $\Theta_{k,S}$, $k = 1, ..., l$.

2 - From a density φ with support Ω, simulate $\{Y_j : j = 1, ..., m_C\}$ independent identically distributed as φ.

3 - Define the objects in (3.14).

4 - Solve the maxmin linear problem:

$$\max_{0 \le x,\ Ax \le b} \ \min_{i=1,..n} x' c_i, \tag{3.15}$$

using a standard maxmin algorithm, for example the *fminimax* function in Matlab.

3.4.2 Minimax Regret Algorithm

Now, consider the n-GMR problem:

$$\inf_{\phi \in \Gamma^n(g,\gamma)} \ \max_{i=1,...,n} \left(\beta^*(\theta^n_{R,i}) - \int_\Omega \phi f_{\theta^n_{R,i}} d\mu \right). \tag{3.16}$$

We use the first algorithm to calculate the general power envelope $d_i := \beta^*(\theta^n_{R,i})$, for $i = 1, ..., n$. Using importance sampling, we can approximate the integrals in the problem (3.16) to obtain the finite minimax regret linear problem:

$$\min_{0 \le x,\ Ax \le b} \ \max_{i=1,...,n} \ (d_i - c_i' x). \tag{3.17}$$

The problem (3.17) can be solved by a standard minimax algorithm, for example the *fminimax* function in Matlab. In the following algorithm we gather all the optimization steps.

General Minimax Regret Algorithm: Solving the n-GMRP.

1 - Let $n \in \mathbb{N}$ be a large integer. For $i = 1, ..., n$, choose:

$$\theta^n_{R,i} \in \Pi^n_{R,i} \text{ and } \theta^n_{k,S,i} \in \Pi^n_{k,S,i} \text{ for all } k = 1, ..., l,$$

where $\{\Pi^n_{R,i}\}^n_{i=1}$ and $\{\Pi^n_{k,S,i}\}^n_{i=1}$ partitions of Θ_R and $\Theta_{k,S}$, $k = 1, ..., l$.

2 - From a density φ with support Ω, simulate $\{Y_j : j = 1, ..., m_C\}$ independent identically distributed as φ.

3 - Calculate $d_i = \beta^*(\theta^n_{R,i})$ for $i = 1, .., n$, using the maxmin algorithm (3.4.1).

4 - Define the objects in (3.14).

5 - Solve the problem:

$$\min_{0 \le x,\ Ax \le b} \ \max_{i=1,...,n} \ (d_i - x' c_i),$$

using a standard minimax algorithm, for example the *fminimax* function in Matlab.

Example 2 *Moment Inequality - Continuing*

Now we illustrate the use GM and GMR tests in the moment inequality testing problem (3.3) with $k = 2$. In the following, we implement the algorithm with φ at (3.11) being a normal density function, number of simulations equals to $m_C = 150,000$ and size equals to $\alpha = 0.05$. Since any similar test ϕ has power $E_\theta[\phi] = \alpha$ for any $\theta \in \{(\theta_1, \theta_2) : \theta_1 = 0, \theta_2 \in \mathbb{R}\} \cup \{(\theta_1, \theta_2) : \theta_1 \in \mathbb{R}, \theta_2 = 0\}$, we should not impose similar or unbiased constraints in the general set $\Gamma(g, \gamma)$. Therefore, we imposed the following contraints to the Maxmin and Minimax Regret tests: i) size constraints for specific points in the null hypothesis;

and ii) power control over some points in the alternative hypothesis. More specifically, we set the following partitions $\Pi_{1,R} = \Pi^n_{R,i} = \{(-3,-3),(-3,-1.5),(-3,0),(-2.25,-3),(-2.25,-1.5),(-2.25,0),(-1.5,-3),$ $(-1.5,-1.5),(-1,0),(0,-3),(0,-2.25),(0,-1.5)\}$ and $\Pi_{1,S} = \Pi^n_{k,S,i} = \{(0,0),(0,0.5),(0,1),(0,1.5),(0,2),$ $(0.5,0),(0.5,0.5),(0.5,1),(0.5,1.5),(0.5,2),(1,0),(1,0.5),(1,1),(1,1.5),(1,2),(1.5,0),(1.5,0.5),(1.5,1),(1.5,1.5),$ $(1.5,2),(2,0),(2,0.5),(2,1),(2,1.5),(2,2)\}$. In order to compare both tests, we derive the power envelope from the unfeasible point optimal test, i.e. we use the likelihood ratio test for each point in the alternative hypothesis to construct the highest power that a test can achieve.

From figure 3.1 to 3.5, we note that both Maxmin and Minimax Regret tests defined by $\Pi_{1,R}$ and $\Pi_{1,S}$ have size α over the null hypothesis $H : \theta \geq 0$. Moreover, both tests have power larger than α at some points $\theta \in \{(\theta_1,\theta_2) : \theta_1 = 0, \theta_2 \in \mathbb{R}\} \cup \{(\theta_1,\theta_2) : \theta_1 \in \mathbb{R}, \theta_2 = 0\}$, hence GM and GMR are nonsimilar (and biased) tests. Note that there is no dominance (in admissible sense) of one test over the other (see, for example, figure 3.1), since the Maxmin test has higher power than the Minimax Regret test at some points and Minimax Regret test has higher power for other alternative points. Overall, Maxmin and Minimax Regret tests have good power compared to the power envelope. For $\theta_2 = 0$, the power envelope and the Minimax Regret and Maxmin powers are 0.90, 0.85 and 0.80 when $\theta_1 = -3$; 0.64, 0.55 and 0.48.when $\theta_1 = -2$, respecively. Interestingly, the power of the tests are still higher for points closer to the null hypothesis. Consider $\theta_2 = -0.5$, for example, the power envelope and the Maxmin and Minimax Regret powers are 0.91, 0.88 and 0.87 when $\theta_1 = -3$; and 0.66, 0.61 and 0.57 when $\theta_1 = -2$.

Now we compare the Maxmin and Minimax Regret tests to the existing test proposed by Andrews and Barwick (2012) (hereinafter, RMS test). According to Andrews (2012), the RMS test has good performance. Table 3.1 presents the Minimax Regret, Maxmin and RMS powers. Overall, there is no strict dominance of one test (either Maxmin, Minimax Regret or RMS) over the others and all tests have performances very similar. For example, Minimax Regret, Maxmin and RMS powers are 0.55, 0.48 and 0.53 at $(\theta_1,\theta_2) = (-2,0)$. At $(\theta_1,\theta_2) = (0.5,-0.5)$, however, the Minimax Regret power is 0.04 when Maxmin and RMS powers are 0.08. Since there is no better test and all tests have good power properties we recommend the use of all three tests.

The results presented in this section highlight the flexibility of the use of the General Maxmin and General Minimax Regret tests. We presented a simple application, with an easy to implement constraints, but the applicability of the methods presented on this thesis are extensive. Maxmin and Minimax Regrets tests can be implemented to any econometric model and with a broader set of contraints, given the assumptions of the previous sections. Furthermore, the theoretical and numerical results for hypothesis testing problems of Chapter 2 and 3 can be easily extended to general decision rule problems.

Figure 3.1: Power envelope and power of the Maxmin and Minimax Regret tests

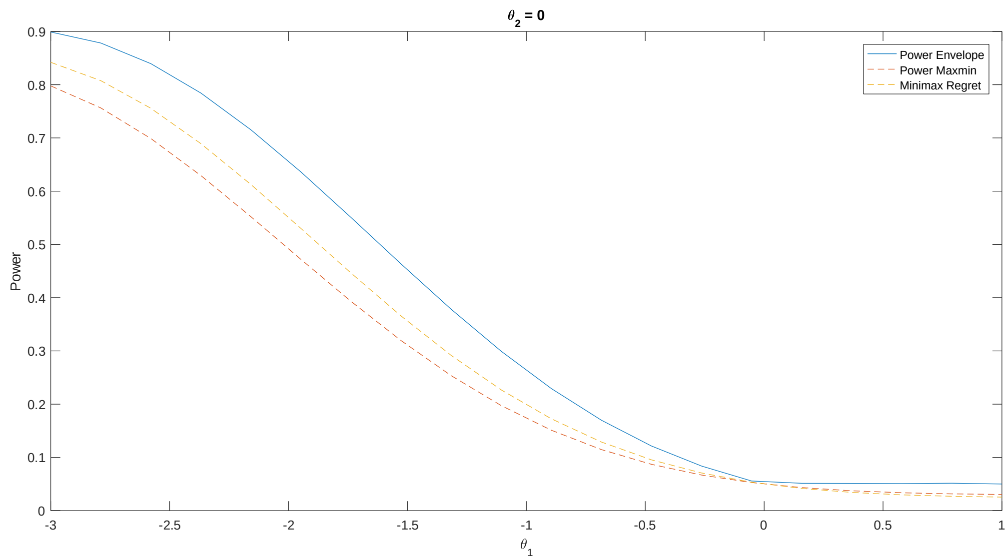

Figure 3.2: Power envelope and power of the Maxmin and Minimax Regret tests

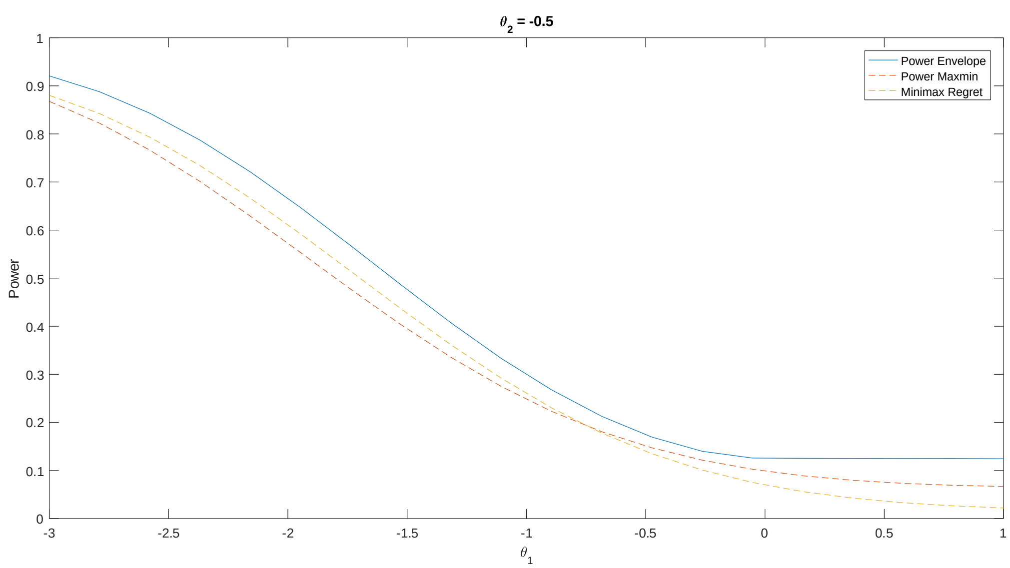

Figure 3.3: Power envelope and power of the Maxmin and Minimax Regret tests

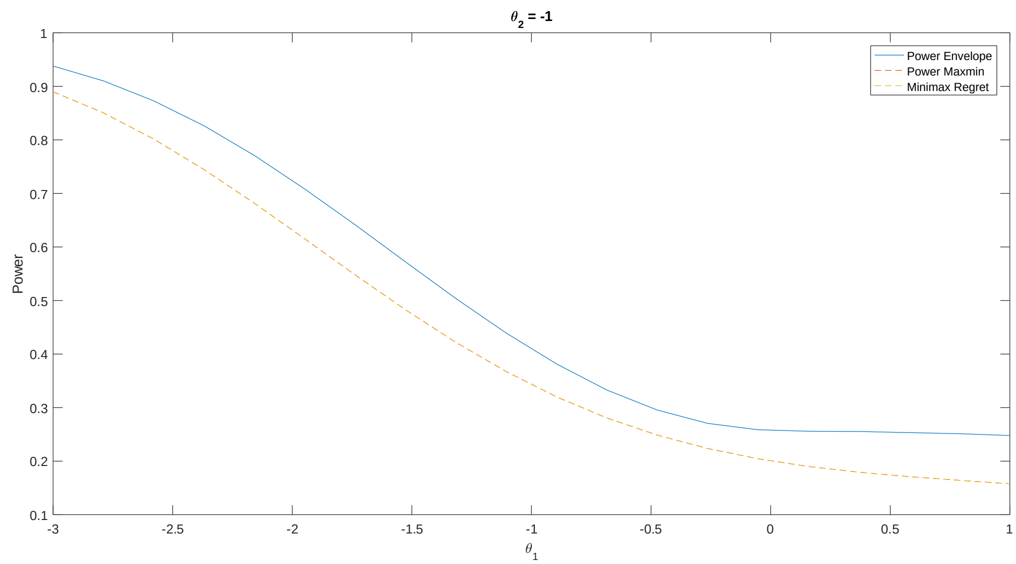

Figure 3.4: Power envelope and power of the Maxmin and Minimax Regret tests

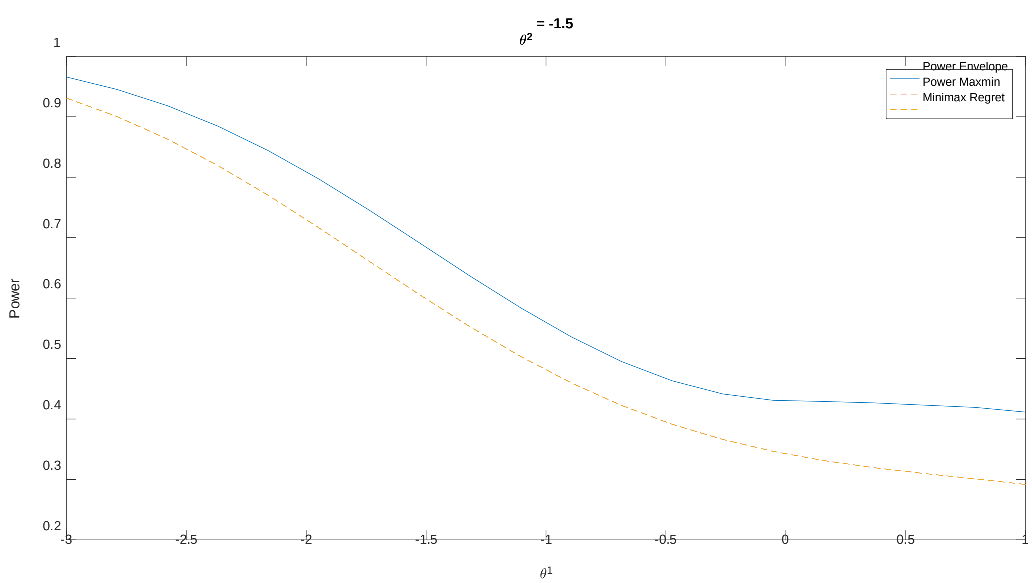

Figure 3.5: Power envelope and power of the Maxmin and Minimax Regret tests

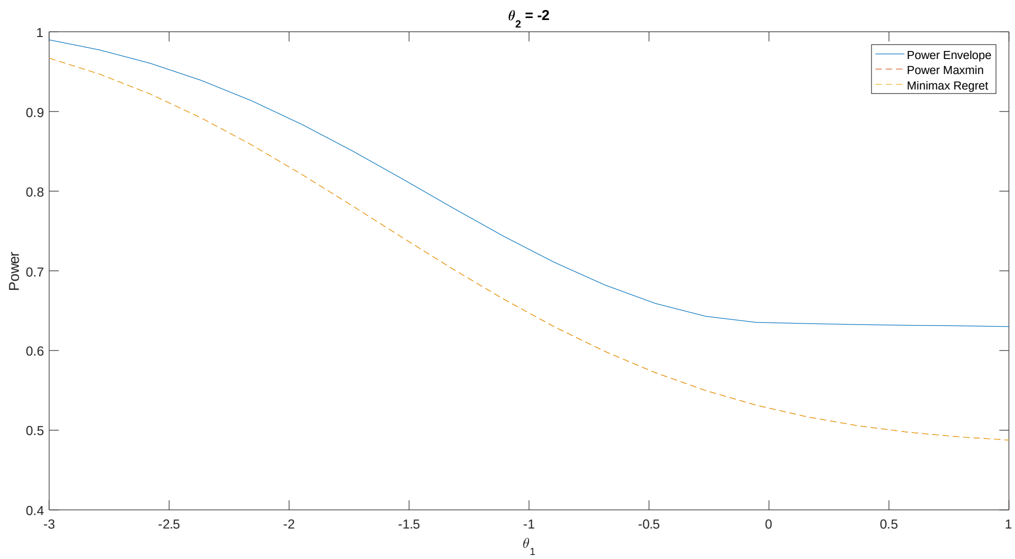

Table 3.1: Power for the Minimax Regret, Maxmin and Andrews–Barwick RMS tests

Minimax Regret			θ_2		
θ_1	0	-0.5	-1	-1.5	-2
1	0.03	0.02	0.16	0.29	0.49
0.5	0.03	0.04	0.18	0.33	0.50
0	0.05	0.08	0.20	0.34	0.54
-0.5	0.10	0.15	0.25	0.40	0.57
-1	0.20	0.25	0.34	0.49	0.65
-2	0.55	0.61	0.63	0.74	0.83
-3	0.85	0.88	0.89	0.93	0.97
Maxmin			θ_2		
θ_1	0	-0.5	-1	-1.5	-2
1	0.03	0.07	0.16	0.29	0.49
0.5	0.03	0.08	0.18	0.33	0.50
0	0.05	0.10	0.20	0.34	0.54
-0.5	0.09	0.16	0.25	0.40	0.57
-1	0.18	0.24	0.34	0.49	0.65
-2	0.48	0.57	0.63	0.74	0.83
-3	0.80	0.87	0.89	0.93	0.97
Andrews-Barwick RMS			θ_2		
θ_1	0	-0.5	-1	-1.5	-2
1	0.03	0.08	0.18	0.33	0.52
0.5	0.03	0.08	0.18	0.33	0.51
0	0.05	0.10	0.20	0.34	0.53
-0.5	0.10	0.15	0.25	0.40	0.57
-1	0.20	0.25	0.35	0.49	0.65
-2	0.53	0.57	0.65	0.74	0.83
-3	0.85	0.87	0.90	0.93	0.96

Chapter 4

Appendix - Proofs

4.1 Proofs of Chapter 1

Derivation of the t-Statistics

The k-class estimator for the model (1.7) is given by:

$$\beta(\underline{k}) = \frac{[y_2^{\perp\prime}(I - \underline{k}M_Z)y_1^{\perp}]}{[y_2^{\perp\prime}(I - \underline{k}M_Z)y_2^{\perp}]} \tag{4.1}$$

where the four $\underline{k}$-class used are:

$$\begin{array}{ll} \text{2SLS:} & \underline{k} = 1, \\ \text{LIMLK:} & \underline{k} = \underline{k}_{\ LIMLK} = \text{Smallest root } \kappa \text{ of } |(Y'P_ZY/n + \Omega) - \kappa\Omega| = 0 \\ \text{B2SLS:} & \underline{k} = n/\left(n - k + 2\right), \\ \text{Fuller:} & \underline{k} = \underline{k}_{\ LIMLK} - 1/\left(n - k - p\right). \end{array} \tag{4.2}$$

Consequently the t-statistics follows:

$$\begin{aligned} t(\underline{k}) &= \frac{\beta(\underline{k}) - \beta_0}{\sigma_u(\underline{k})[y_2^{\perp\prime}(I - \underline{k}M_Z)y_2^{\perp}]^{-1/2}}, \text{ where} \\ \sigma_u^2(\underline{k}) &= b(\underline{k})'\Omega b(\underline{k}) \text{ and } b(\underline{k}) = (1, -\beta(\underline{k}))' \end{aligned} \tag{4.3}$$

Since we known Ω we can simplify some expressions:

$$\begin{aligned} [y_2^{\perp\prime}(I - \underline{k}M_Z)y_l^{\perp}] &= [y_2^{\perp\prime}P_Zy_l^{\perp} - (\underline{k} - 1)y_2^{\perp\prime}M_Zy_l^{\perp}] \\ &= [y_2P_Zy_l - n(\underline{k} - 1)(n - k - p)/n \cdot e_2'Y'M_XM_ZYe_l/(n - k - p)] \\ &= [y_2P_Zy_l - n(\underline{k} - 1)(n - k - p)/n \cdot \widehat{w}_{2l}] \\ &= [y_2P_Zy_l - n(\underline{k} - 1)w_{2l}] + o_p(1) \end{aligned} \tag{4.4}$$

where $e_1 = (1, 0)'$, $e_2 = (0, 1)'$ and $l = 1, 2$, the last equality follows from the fact that $\hat{\Omega}$ is a consistent estimator, $(n - k - p)/n - 1 = o(1)$ and $n(\underline{k} - 1) = O_p(1)$.

So the t-statistic can be simplified too:

$$\begin{aligned} t(\underline{k}) &= \frac{\beta(\underline{k}) - \beta_0}{\sigma_u(\underline{k})[y_2'P_Zy_2 - n(\underline{k} - 1)\omega_{22}]^{-1/2}}, \text{ where} \\ \beta(\underline{k}) &= \frac{[y_2'P_Zy_1 - n(\underline{k} - 1)\omega_{21}]}{[y_2'P_Zy_2 - n(\underline{k} - 1)\omega_{22}]}, \end{aligned} \tag{4.5}$$

$$\sigma_u^2(\underline{k}) = b(\underline{k})'\Omega b(\underline{k}) \text{ and } b(\underline{k}) = (1, -\beta(\underline{k}))'. \tag{4.6}$$

Derivation of the One-Sided Likelihood Ratio Statistics

Ignoring an additive constant, the log-likelihood function for known Ω with all parameters concentrated out except β is

$$l_c(Y;\beta,\Omega) = -\frac{n}{2}\ln\det(\Omega) - \frac{1}{2}\left(tr(\Omega^{-1}Y'M_XY) + R(\beta)\right). \tag{4.7}$$

Hence, we have

$$LR1 = 2\left[\sup_{\beta\geq\beta_0} l_c(Y;\beta,\Omega) - l_c(Y;\beta_0,\Omega)\right] = R(\beta_0) - \inf_{\beta\geq\beta_0} R(\beta). \tag{4.8}$$

We now determine $\inf_{\beta\geq\beta_0} R(\beta)$. By definition, $\beta(\underline{k}_{LIMLK})$ maximizes $l_c(Y;\beta,\Omega)$ over $\beta\in\mathbb{R}$. Equivalently, $\beta(\underline{k}_{LIMLK})$ minimizes $R(\beta)$ over $\beta\in\mathbb{R}$. If $\beta(\underline{k}_{LIMLK})\geq\beta_0$, then $\inf_{\beta\geq\beta_0} R(\beta) = \inf_{\beta\in\mathbb{R}} R(\beta) = R(\beta(\underline{k}_{LIMLK}))$ and $LR1 = R(\beta_0) - \inf_{\beta\in\mathbb{R}} R(\beta) = LR$. If $\beta(\underline{k}_{LIMLK}) < \beta_0$, then $\inf_{\beta\geq\beta_0} R(\beta)$ equals either $R(\beta_0)$ or $R(\infty)$ because $R(\beta)$ is the ratio of two quadratic forms in β with positive definite weight matrices. Hence, the second equality in (1.21) holds. A similar reasoning yields (1.23).

We now provide an expression for $\beta(\underline{k}_{LIMLK})$. The LIMLK estimator maximizes $l_c(Y;\beta,\Omega)$ or minimizes $R(\beta)$. Note that:

$$R(\beta) = \frac{\widetilde{b}'JQJ\widetilde{b}}{\widetilde{b}'\widetilde{b}}, \text{ where } \widetilde{b} = \Omega^{1//2}b. \tag{4.9}$$

The minimum of $R(\beta)$ is obtained by the eigenvector $\widetilde{b}^*$ that corresponds to the smallest eigenvalue of JQJ. Hence,

$$\beta(\underline{k}_{LIMLK}) = -b_2^*/b_1^*, \text{ where } b^* = \Omega^{-1//2}\widetilde{b}^*.$$

Proof. [Theorem **1**] The power function is given by

$$K(\phi;\beta,\lambda) = \int_{\mathbb{R}^+\times\mathbb{R}\times\mathbb{R}^+} \phi(q_S,q_{ST},q_T) f_Q(q_S,q_{ST},q_T;\beta,\lambda)dq_S dq_{ST} dq_T. \tag{4.10}$$

We want to find a test that maximizes power at (β^*,λ^*) among all level α invariant similar tests. By Theorem 2 of AMS06a, invariant similar tests must be similar conditional on $Q_T = q_T$ for almost all q_T. In addition, the unconditional power equals the expected conditional power given Q_T. Hence, it is sufficient to determine the test that maximizes conditional power given $Q_T = q_T$ among invariant tests that are similar conditional on $Q_T = q_T$, for each q_T. By the Neyman-Pearson Lemma, the test of significance level α that maximizes conditional power given $Q_T = q_T$ is of the likelihood ratio (LR) form and rejects H_0 when the LR is sufficiently large (part a) or small (part b). In particular, the conditional LR test statistic is

$$LR_{\beta^*\lambda^*}(q_S,q_{ST},q_T) = \frac{f_{Q_S,Q_{ST}|Q_T}(q_S,q_{ST}|q_T;\beta^*,\lambda^*)}{f_{Q_S,Q_{ST}|Q_T}(q_S,q_{ST}|q_T;\beta_0)} = \frac{f_Q(q;\beta^*,\lambda^*)}{f_{Q_T}(q_T;\beta^*,\lambda^*)f_{Q_S,Q_{ST}|Q_T}(q_S,q_{ST}|q_T;\beta_0)}. \tag{4.11}$$

From the density $f_Q(q;\beta,\lambda)$ given in (1.13), we can determine $f_{Q_T}(q_T;\beta^*,\lambda^*)$ and $f_{Q_1|Q_T}(q_1|q_T;\beta_0)$ to provide the explicit expression for $LR_{\beta^*\lambda^*}(Q_1,Q_T)$ that appears in (1.30); see Lemma 3 of AMS06a. ■

Proof. [Theorem **3**] First, we rewrite $LR1$ in a form that is closer to that of $\widetilde{\xi}_{\widehat{\beta}}$. Ignoring an additive constant, the log-likelihood function (after concentrating out η) can be written as

$$l(Y;\beta,\pi,\Omega) = -\frac{n}{2}\ln|\Omega| - \frac{1}{2}tr\left(\Omega^{-1}V'M_XV\right), \text{ where } V = Y - Z\pi a' - X\eta. \tag{4.12}$$

Maximizing (4.12) with respect to π, one finds that $\pi(\beta) = (Z'Z)^{-1} Z'Y\Omega^{-1}a/a'\Omega^{-1}a$, where $a\equiv(\beta\ ,1)'$. The concentrated log-likelihood function, $l_c(Y;\beta,\Omega)$, defined as $l(Y;\beta,\pi(\beta),\Omega)$, is given by

$$l_c(Y;\beta,\Omega) = -\frac{n}{2}\ln|\Omega| - \frac{1}{2}\left[tr(\Omega^{-1}Y'M_{\overline{Z}}Y) - \frac{a'\Omega^{-1}Y'P_ZY\Omega^{-1}a}{a'\Omega^{-1}a}\right]. \tag{4.13}$$

We can simplifies the last term in (4.13):

$$\begin{aligned}\tau(Q;\beta,\Omega) &= \frac{a'\Omega^{-1}Y'P_ZY\Omega^{-1}a}{a'\Omega^{-1}a} \\ &= \frac{a'\Omega^{-1/2}J(J'\Omega^{-1/2}Y'P_ZY\Omega^{-1/2}J)J'\Omega^{-1/2}a}{a'\Omega^{-1/2}JJ'\Omega^{-1/2}a} \\ &= \frac{\widetilde{a}'Q\widetilde{a}}{\widetilde{a}'\widetilde{a}},\ \text{where } \widetilde{a} = J\Omega^{-1/2}a.\end{aligned} \tag{4.14}$$

When evaluated at $\widehat{\beta}$, the maximum likelihood estimator of β under $H_1 : \beta > \beta_0$, (4.13) becomes

$$l_c(Y;\widehat{\beta},\Omega) = -\frac{n}{2}\ln|\Omega| - \frac{1}{2}\left[tr(\Omega^{-1}Y'Y) - \tau(Q;\widehat{\beta},\Omega)\right]. \tag{4.15}$$

Because $LR1$ is defined as $2[l_c(Y;\widehat{\beta},\Omega) - l_c(Y;\beta_0,\Omega)]$, it follows that

$$LR1 = 2[l_c(Y;\widehat{\beta},\Omega) - l_c(Y;\beta_0,\Omega)] = \tau(Q;\widehat{\beta},\Omega) - Q_T. \tag{4.16}$$

Since the term Q_T can be ignored for conditional testing, the CLR1 test is equivalent to rejecting H_0 when $\tau(Q;\widehat{\beta},\Omega) > \kappa_{\tau,\alpha}(Q_T)$. Because the vector $\widehat{a} = (\widehat{\beta},1)'$ maximizes (4.13),

$$\tau(Q;\widehat{\beta},\Omega) = x'_{\widehat{\beta}}Qx_{\widehat{\beta}},\ \text{where } x_{\widehat{\beta}} = (c_{\widehat{\beta}}/||h_{\widehat{\beta}}||, d_{\widehat{\beta}}/||h_{\widehat{\beta}}||)'. \tag{4.17}$$

This proves the equivalence between the CLR1 test and the empirical POIS test based on $\widetilde{\xi}_{\widehat{\beta}} = x'_{\widehat{\beta}}Qx_{\widehat{\beta}}$. ■

Proof. [Theorem 4] Each test statistic is a continuous function of $\widehat{\Omega}_n$, $\widehat{Q}_{S,n}$, $\widehat{Q}_{ST,n}$ and $\widehat{Q}_{T,n}$ almost everywhere (a.e.). By the Continuous Mapping Theorem, the test statistics have limiting distribution that depends on Ω, $Q_{S,\infty}$, $Q_{ST,\infty}$ and $Q_{T,\infty}$ The same argument applies to the critical value functions. Hence, a) is proved.

Part b) follows from a).

For part c) the test statistics can be written as functions of $\widehat{\Omega}_n$, $\widehat{Q}_{(k-1),n}$, $\widehat{LM1}_n$ and $\widehat{Q}_{T,n}$ (see Sections 3 and 4 of MMV). By the Continuous Mapping Theorem, their asymptotic distributions depends on Ω, $Q_{(k-1),\infty}$, $LM1_\infty$ and $Q_{T,\infty}$. The statistics ($Q_{(k-1),\infty}$, $LM1_\infty$) are independent from $Q_{T,\infty}$ under $\beta = \beta_0$. Their limiting test statistics conditional on $Q_{T,\infty} = q_{T,\infty}$ have parameter-free distributions. The term

$$[y_2'P_Zy_2 + n(1-\underline{k}_{2SLS})\omega_{22}]^{1/2} = [y_2'P_Zy_2]^{1/2}$$

is positive (with probability one) for the 2SLS estimator. The term

$$[y_2'P_Zy_2 + n(1-\underline{k}_{LIML})\omega_{22}]^{1/2}$$

is also nonnegative for the LIML estimator because

$$\frac{y_2'P_Zy_2}{\omega_{22}} \geq \min_b \frac{b'Y'P_ZYb}{b'\Omega b}$$

with probability one. The term

$$[y_2'P_Zy_2 + n(1-\underline{k}_{FULL})\omega_{22}]^{1/2} = [y_2'P_Zy_2 + n(1-\underline{k}_{LIML})\omega_{22} + \omega_{22}]^{1/2}$$

for the Fuller estimator is always positive. The conditional distributions of the t-statistics for the 2SLS, LIML and Fuller estimators are continuous under $\beta = \beta_0$. Hence, the respective conditional t-test are similar. The rejection probability at $\beta = \beta_0$ of the conditional B2SLS t-tests is smaller or equal than α (by the definition of the critical value function). The conditional test based on $LM1$ is conditonally similar and by the law of iterated expectations we establish the result. The tests based on LR1 and MLR1 can be written as function of the independent statistics $Q_{(k-1)}, LM1$ and Q_T (when $\beta = \beta_0$), which guarantees that the conditional distribution is continuous for any q_T (provided the significance level $\alpha \in (0, 1/2)$). ■

Proof. [Lemma 5] For part (a), we need to analyze t-statistics based on the 2SLS, B2SLS, LIMLK and Fuller estimators. The null distribution of the t-statistics conditional on $Q_T = q_T$ depends on the null distribution of Q_S and Q_{ST}. After some calculation, the t-statistic can be written as:

$$\begin{aligned} t\left(\underline{k}\right) &= \frac{\beta\left(\underline{k}\right)-\beta_0}{\sigma_u\left(\underline{k}\right)\left[c_{21}^2 Q_S + 2c_{21}c_{22}Q_{ST} + c_{22}^2 Q_T + n\left(1-\underline{k}\right)\omega_{22}\right]^{-1/2}}, \text{ where} \\ \beta\left(\underline{k}\right) &= \frac{c_{11}c_{21}Q_S + (c_{12}c_{21}+c_{11}c_{22})Q_{ST} + c_{12}c_{22}Q_T + n\left(1-\underline{k}\right)\omega_{12}}{c_{21}^2 Q_S + 2c_{21}c_{22}Q_{ST} + c_{22}^2 Q_T + n\left(1-\underline{k}\right)\omega_{22}}. \end{aligned} \tag{4.18}$$

For the 2SLS estimator, the null conditional distribution on $Q_T = q_T$ of the t-statistic is:

$$\begin{aligned} t(1) &= \frac{\left(c_{21}\frac{Q_S}{q_T^{1/2}} + c_{22}\mathcal{S}_2 Q_S^{1/2}\right)}{\left(c_{22}^2 + c_{21}^2\frac{Q_S}{q_T} + 2c_{21}c_{22}\mathcal{S}_2\frac{Q_S^{1/2}}{q_T^{1/2}}\right)^{1/2}} \cdot \frac{\sigma_u}{\sigma_u(1)} \\ &\rightarrow_d \mathcal{S}_2 Q_S^{1/2} = LM1 \text{ as } q_T \rightarrow_p \infty \end{aligned} \tag{4.19}$$

because $\beta\left(1\right) \rightarrow_p \beta_0$ and $\sigma_u\left(1\right) \rightarrow_p (b_0'\Omega b_0)^{1/2} = \sigma_0$ as $q_T \rightarrow_p \infty$. The same limiting result holds for the t-statitistic based on the B2SLS estimator. Now lets analyse the t-statistic based on the LIMLK estimator. First, by expression (A.12) of AMS06a, the null conditional distribution of the LR statistics converges to chi-square-one as $q_T \rightarrow_p \infty$:

$$\begin{aligned} LR &= \frac{1}{2}\left[Q_S - q_T + (q_T - Q_S)\left(1 + \frac{2q_T}{(q_T-Q_S)^2}Q_S\mathcal{S}_2^2\right)\right] + o_p\left(1\right) \\ &= Q_S\mathcal{S}_2^2 + o_p\left(1\right). \end{aligned} \tag{4.20}$$

and then $n(\underline{k}_{LIMLK}-1)/q_T^{1/2} \rightarrow_p 0$ when $q_T \rightarrow_p \infty$. Second, the null conditional distribution of the LIMLK estimator of β converges in probability to β_0 as $q_T \rightarrow_p \infty$:

$$\beta(\underline{k}_{LIMLK}) \rightarrow_p c_{12}/c_{22} = \beta_0 \tag{4.21}$$

consequently $\sigma_u\left(\underline{k}_{LIMLK}\right) \rightarrow_p \sigma_0$ when $q_T \rightarrow_p \infty$. Finally the null conditional distribution of $t\left(\underline{k}_{LIMLK}\right)$ converges in distribution to standard normal distribution:

$$\begin{aligned} t(\underline{k}_{LIMLK}) &= \frac{(\beta(\underline{k}_{LIMLK})-\beta_0)}{\sigma_u\left(\underline{k}_{LIMLK}\right)[y_2'P_Z y_2 - \kappa_{LIMLK}\cdot w_{22}]^{-1/2}} \\ &\rightarrow_d LM1 \text{ as } q_T \rightarrow_p \infty \end{aligned} \tag{4.22}$$

We can obtain the same result for the t-statistic based on the Fuller estimator. For part (b), note first that the null conditional distribution of $\max\{R\left(\beta_0\right) - R\left(\infty\right), 0\}$ goes to zero in probability as $q_T \rightarrow_p \infty$:

$$\max\left\{Q_S(w_{22}-f_2^2) - g_2 q_T^{1/2}\left(\frac{g_2 q_T^{1/2} + 2f_2\mathcal{S}_2 Q_S^{1/2}}{w_{22}}\right), 0\right\} \rightarrow_p 0. \tag{4.23}$$

and so we can aply the continuous mapping theorem to obtain:

$$\begin{aligned} \sqrt{LR1} &= \sqrt{LR}\times 1(t\left(\underline{k}_{\ LIMLK}\right) > 0) + o_p\left(1\right) \\ &= \sqrt{\mathcal{S}_2^2 Q_S}\times 1(\mathcal{S}_2 Q_S^{1/2} > 0) + o_p\left(1\right) \\ &\rightarrow_d \max\{\mathcal{S}_2 Q_S^{1/2}, 0\} \text{ as } q_T \rightarrow \infty. \end{aligned} \tag{4.24}$$

The critical value for $\max\{Q_S^{1/2}\mathcal{S}_2, 0\}$ at level α (with $0 < \alpha < 1/2$) is z_α because

$$P\left(\max\{Q_S^{1/2}\mathcal{S}_2, 0\} \geq z_\alpha\right) = P\left(Q_S^{1/2}\mathcal{S}_2 \geq z_\alpha\right) = \alpha. \tag{4.25}$$

Part (c) also follows from (4.24) and (4.25) and (4.23). ■

Proof. [Theorem 6] For part (a),

$$\begin{aligned} Y'P_ZY + n\left(1-\underline{k}\right)\widehat{\Omega}_n &= Y'P_ZY + n\left(1-\underline{k}\right)\Omega + o_p\left(1\right) \text{ and} \\ \widehat{\sigma}_u^2\left(\underline{k}\right) &= \sigma_u^2\left(\underline{k}\right) + o_p\left(1\right) \ \left(= \sigma_0^2 + o_p\left(1\right) \text{ as well}\right), \end{aligned} \tag{4.26}$$

if $\underline{k} - 1 = O_p\left(n^{-1}\right)$ and $\widehat{\Omega}_n$ is a consistent estimator of Ω. Hence, $\widehat{t}\left(\underline{k}\right) = t\left(\underline{k}\right) + o_p\left(1\right)$. The $t\left(\underline{k}\right)$ statistic in turn is given by

$$t\left(\underline{k}\right) = \frac{n^{-1/2}y_2'P_Z\left(y_1 - y_2\beta_0\right) + n^{1/2}\left(1-\underline{k}\right)\left(\omega_{12} - \omega_{22}\beta_0\right)}{\sigma_u\left(\underline{k}\right)\left[n^{-1}y_2'\,P_Z\,y_2 + \left(1-\underline{k}\right)\omega_{22}\right]^{1/2}}. \tag{4.27}$$

Because $\underline{k} - 1 = O_p\left(n^{-1}\right)$, we have

$$\begin{aligned} & n^{-1/2}y_2'P_Z\left(y_1 - y_2\beta_0\right) + n^{1/2}\left(1-\underline{k}\right)\left(\omega_{12} - \omega_{22}\beta_0\right) \\ &= n^{-1/2}y_2'P_ZYb_0 + o_p(1) = \alpha_T'S\left(b_0'\Omega b_0/a_0'\Omega^{-1}a_0\right)^{1/2} + o_p\left(1\right). \end{aligned} \tag{4.28}$$

Analogously,

$$n^{-1}y_2'\,P_Z\,y_2 + \left(1-\underline{k}\right)\omega_{22} = \pi'D_Z\pi + o_p\left(1\right) = \alpha_T'\alpha_T/(a_0'\Omega^{-1}a_0) + o_p\left(1\right). \tag{4.29}$$

Using the fact that $\sigma_u\left(\underline{k}\right) \rightarrow_p \left(b_0'\Omega b_0\right)^{1/2}$, we have

$$t\left(\underline{k}\right) \rightarrow_d \left(\alpha_T'S_{B\infty}\right)/||\alpha_T||. \tag{4.30}$$

Finally, using $\widehat{\underline{k}}$ instead of $\underline{k}$ does not have any effect asymptotically. The proof of part (b) follows immediately from (1.54). For part (c), recall that the $LR1$ statistic is

$$LR1 = LR \times 1(\beta\left(\underline{k}_{\ LIMLK}\right) > \beta_0) + \max\left\{R(\beta_0) - R\left(\infty\right), 0\right\} \times 1(\beta\left(\underline{k}_{\ LIMLK}\right) < \beta_0). \tag{4.31}$$

Under local alternatives $\beta = \beta_0 + B/n^{1/2}$, $\max\left\{R(\beta_0) - R\left(\infty\right), 0\right\} \rightarrow_p 0$ because $R(\beta_0)$ is $O_p\left(1\right)$ and $R\left(\infty\right) \rightarrow_p \infty$. Therefore,

$$\begin{aligned} LR1 &= LR \times 1(t\left(\underline{k}_{\ LIMLK}\right) > 0) + o_p\left(1\right) \\ &\rightarrow_d \left(\alpha_T'S_{B\infty}\right)^2/||\alpha_T||^2 \times 1[\alpha_T'S_{B\infty}/||\alpha_T|| > 0], \end{aligned} \tag{4.32}$$

where the third equality follows from the continuous mapping theorem and the joint convergence of LR and $t\left(\underline{k}_{\ LIMLK}\right)$ to $(\alpha_T'S_{B\infty})^2/||\alpha_T||^2$ and $\alpha_T'S_{B\infty}/||\alpha_T||$, respectively; see AMS06a, Theorem 6(c), regarding the convergence in distribution of LR to $(\alpha_T'S_{B\infty})^2/||\alpha_T||^2$. Part (d) also follows from (4.32) because $\max\{R\left(\beta_0\right) - R\left(\infty\right), 0\}$ converges in probability to zero. ■

Proof. [Theorem 7] Following the proof of Theorem 7 of AMS06a, we know that the one-sided LM statistic for known Ω is $LM1_n = Q_{ST,n}/Q_{T,n}^{1/2}$, which is asymptotically efficient by standard results. By Lemma 5, the critical values of conditional tests based on $LR1_n$, $MLR1_n$, and t-statistics converge to a standard normal $1-\alpha$ quantile (provided $\alpha \in [0, 1/2)$ for the likelihood ratio statistics). The $LR1_n^{1/2}$ and $MLR1_n^{1/2}$ statistics are not asymptotically equivalent to $LM1_n$. However, the asymptotic power of the one-sided tests based on $LR1_n^{1/2}$, $MLR1_n^{1/2}$, and $LM1_n$ are the same:

$$\begin{aligned} P\left(\max\{(\alpha_T'S_{B\infty})/||\alpha_T||, 0\} \geq z_\alpha\right) &= P(\max\{\varsigma_1 + \lambda^{1/2}B\left(b_0'\Omega b_0\right)^{-1/2}, 0\} \geq z_\alpha) \\ &= P(\varsigma_1 + \lambda^{1/2}B\left(b_0'\Omega b_0\right)^{-1/2} \geq z_\alpha) \\ &= P\left(\alpha_T'S_{B\infty}/||\alpha_T|| \geq z_\alpha\right), \end{aligned} \tag{4.33}$$

where $\varsigma_1 \sim N(0,1)$, $B > 0$, and z_α is a positive critical value. By Theorem 6, the asymptotic behavior of the tests above are the same when $\widehat{\Omega}_n$ replaces Ω. Hence, these tests are asymptotically efficient when Ω is estimated. ■

Proof. [Lemma 8] Part (i) of the Lemma is established as follows:

$$\begin{aligned} S_n/n^{1/2} &= (n^{-1}Z'Z)^{-1/2}n^{-1}Z'Yb_0 \cdot (b_0'\Omega b_0)^{-1/2} \\ &\to_p D_Z^{1/2}\pi a'b_0 \cdot (b_0'\Omega b_0)^{-1/2} = D_Z^{1/2}\pi c_\beta. \end{aligned} \tag{4.34}$$

Similarly,

$$\begin{aligned} T_n/n^{1/2} &= (n^{-1}Z'Z)^{-1/2}n^{-1}Z'Y\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2} \\ &\to_p D_Z^{1/2}\pi a'\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2} = D_Z^{1/2}\pi d_\beta. \end{aligned} \tag{4.35}$$

Part (ii) of the Lemma follows from Lemma 1 of AMS06b and part (i).Next, we prove part (iii) of the Lemma. If $\beta = \beta_{AR}$, then $a'\Omega^{-1}a_0 = 0$ and using Assumption 4, we have

$$T_n = (n^{-1}Z'Z)^{-1/2}n^{-1/2}Z'V\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2} \tag{4.36}$$

$$\to_d \varsigma_k \sim N(0, I_k). \tag{4.37}$$

Part (iii) now follows from Lemma 1 of AMS06b. ■

Proof. [Theorem 9] For part (a),

$$\begin{aligned} Y'P_ZY + n\,(1-\underline{k})\,\widehat{\Omega}_n &= Y'P_ZY + n\,(1-\underline{k})\,\Omega + o_p\,(1) \text{ and} \\ \widehat{\sigma}_u^2\,(\underline{k}) &= \sigma_u^2\,(\underline{k}) + o_p\,(1) \;\; (= \sigma_0^2 + o_p\,(1) \text{ as well}), \end{aligned} \tag{4.38}$$

if $\underline{k} - 1 = O_p\left(n^{-1}\right)$ and $\widehat{\Omega}_n$ is a consistent estimator of Ω. Hence, $\widehat{t}\,(\underline{k}) = t\,(\underline{k}) + o_p\,(1)$. The $t\,(\underline{k})$ statistic in turn is given by

$$t\,(\underline{k}) = \frac{n^{-1/2}y_2'P_Z\,(y_1 - y_2\beta_0) + n^{1/2}\,(1-\underline{k})\,(\omega_{12} - \omega_{22}\beta_0)}{\sigma_u\,(\underline{k})\,[n^{-1}y_2'\,P_Z\,y_2 + (1-\underline{k})\,\omega_{22}]^{1/2}}. \tag{4.39}$$

Because $\underline{k} - 1 = O_p\left(n^{-1}\right)$, we have

$$\begin{aligned} & n^{-1/2}y_2'P_Z\,(y_1 - y_2\beta_0) + n^{1/2}\,(1-\underline{k})\,(\omega_{12} - \omega_{22}\beta_0) \\ &= n^{-1/2}y_2'P_ZYb_0 + o_p(1) = \alpha_T'S\left(b_0'\Omega b_0/a_0'\Omega^{-1}a_0\right)^{1/2} + o_p\,(1)\,. \end{aligned} \tag{4.40}$$

Analogously,

$$n^{-1}y_2'\,P_Z\,y_2 + (1-\underline{k})\,\omega_{22} = \pi'D_Z\pi + o_p\,(1) = \alpha_T'\alpha_T/(a_0'\Omega^{-1}a_0) + o_p\,(1)\,. \tag{4.41}$$

Using the fact that $\sigma_u\,(\underline{k}) \to_p (b_0'\Omega b_0)^{1/2}$, we have

$$t\,(\underline{k}) \to_d (\alpha_T'S_{B\infty})/||\alpha_T||. \tag{4.42}$$

Finally, using $\widehat{\underline{k}}$ instead of $\underline{k}$ does not have any effect asymptotically. The proof of parts (b)-(c) follows immediately from (1.54). For parts (d)-(e), recall that the $LR1$ statistic is

$$LR1 = LR \times 1(\beta\,(\underline{k}_{\;LIMLK}) > \beta_0) + \max\{R(\beta_0) - R\,(\infty)\,, 0\} \times 1(\beta\,(\underline{k}_{\;LIMLK}) < \beta_0). \tag{4.43}$$

Under local alternatives $\beta = \beta_0 + B/n^{1/2}$, $\max\{R(\beta_0) - R\,(\infty)\,, 0\} \to_p 0$ because $R(\beta_0)$ is $O_p\,(1)$ and $R\,(\infty) \to_p \infty$. Therefore,

$$\begin{aligned} LR1 &= LR \times 1(t\,(\underline{k}_{\;LIMLK}) > 0) + o_p\,(1) \\ &\to_d (\alpha_T'S_{B\infty})^2/||\alpha_T||^2 \times 1[\alpha_T'S_{B\infty}/||\alpha_T|| > 0], \end{aligned} \tag{4.44}$$

where the third equality follows from the continuous mapping theorem and the joint convergence of LR and $t\,(\underline{k}_{\;LIMLK})$ to $(\alpha_T'S_{B\infty})^2/||\alpha_T||^2$ and $\alpha_T'S_{B\infty}/||\alpha_T||$, respectively; see AMS06a, Theorem 6(c), regarding the convergence in distribution of LR to $(\alpha_T'S_{B\infty})^2/||\alpha_T||^2$. Parts (f)-(g) also follows from (4.32) because $\max\{R\,(\beta_0) - R\,(\infty)\,, 0\}$ converges in probability to zero. ■

Proof. [Proposition 10] In the proofs we use LLN for the Law of Large Numbers for a triangular array of row-wise i.i.d. random variables, CLT for the Lyapunov's triangular array central limit theorem, CMT for Continuous Mapping Theorem and Slutsky for the Slutksy's Theorem. The similar parameter space is

$$\Theta^S = \left\{ \begin{array}{c} \theta = (\theta_1, \theta_2, \theta_{3F}, \theta_4, \theta_{5F}) : \theta_1 = (||\pi||),\ \theta_2 = (\pi/||\pi||) \in \mathbb{R}^k, \\ \theta_{3F} = (E_F(Z_i Z_i'), E_F(V_i V_i')),\ \theta_4 = (\gamma_1, \xi_1, \beta_0) \in \mathbb{R}^{2p+1},\ \theta_{5F} = F, \\ \text{where } E_F(Z_i Z_i') \in \mathbb{R}^{k\times k}_{p.d} \text{ and } E_F(v_i v_i') \in \mathbb{R}^{2\times 2}_{p.d}. \end{array} \right\}$$

To use Corollary 2.1 (c) of Andrews, Cheng, and Guggenberger (2011) we need to prove that the asymptotic rejection probability of the conditional one-sided t-tests equals α under all subsequences $\{p_n\}_{n\in\mathbb{N}}$ of $\{n\}_{n\in\mathbb{N}}$ and all sequences $\{\theta_{p_n} \in \Theta^S : n \in \mathbb{N}\}$ for which

$$h_{p_n}(\theta_{p_n}) = (\sqrt{p_n}||\pi_{p_n}||, \pi_{p_n}/||\pi_{p_n}||, E_{\theta_{p_n}}(Z_i Z_i'), E_{\theta_{p_n}}(V_i V_i')) \to h$$

where $h \in (\mathbb{R} \cup \{+\infty\}) \times \mathbb{R}^k \times \mathbb{R}^{k\times k}_{p.d} \times \mathbb{R}^{2\times 2}_{p.d}$. We prove for the full sequence $\{n\}_{n\in\mathbb{N}}$ since the proof goes through with $\{p_n\}_{n\in\mathbb{N}}$ too. Let $\{\theta_n\}_{n\in\mathbb{N}}$ be a sequence in Θ^S such that $h_n(\theta_n) \to h = (h_1, h_2, h_{31}, h_{32})$ where:

$$h_1 = \lim_{n\to\infty} \sqrt{n}||\pi_n||,\ h_2 = \lim_{n\to\infty} \frac{\pi_n}{||\pi_n||},\ h_{31} = \lim_{n\to\infty} E_{\theta_n}(Z_i Z_i'),\ h_{32} = \lim_{n\to\infty} E_{\theta_n}(V_i V_i').$$

Given the discontinuity in the asymptotic distribution of the t-statistics in θ_1, we separate the proof in two cases: $h_1 < \infty$ and $h_1 = \infty$ (strong instruments). The proof for the case $h_1 < \infty$ follows closely to the proof of Lemma 4 of AMS06a in WIV asymptotics. For $h_1 = \infty$, the arguments used to prove the result are in the same line as Lemma 6 of AMS06a in the SIV asymptotics.

Case 1) $h_1 < \infty$. First note that $\widehat{\Omega}_n \to_p h_{32}$ by the LLN and the moment conditions. The asymptotic behavior of $\widehat{\Omega}_n$ does not depend on whether h_1 is finite or not. Now we analyze the asymptotic distribution of $\widehat{S}_n$:

$$\begin{aligned} \widehat{S}_n &= (b_0'\widehat{\Omega}_n b_0)^{-1/2} \left(b_0' \otimes \left(n^{-1}Z'Z\right)^{-1/2}\right) vec(n^{-1/2}Z'V) \\ &\to_d (b_0' h_{32} b_0)^{-1/2} \left(b_0' \otimes (h_{31})^{-1/2}\right) \times N(0, h_{32} \otimes h_{31}) \\ &= S^h \sim N(0, I_k), \end{aligned} \tag{4.45}$$

where the second equality uses the LLN, CMT, CLT, Slutsky and homoskedastic assumptions. The asymptotic distribution of $\widehat{S}_n$ does not use the fact that $h_1 < \infty$, so it is the same for the case when h_1 is infinity. The asymptotic distribution of $\widehat{T}_n$ depends on $h_1 < \infty$:

$$\begin{aligned} \widehat{T}_n &= (a_0'\widehat{\Omega}_n^{-1} a_0)^{1/2} (n^{-1}Z'Z)^{1/2} n^{1/2} ||\pi_n|| \times \pi_n/||\pi_n|| \\ &\quad + (a_0'\widehat{\Omega}_n^{-1} a_0)^{-1/2} \left(a_0'\widehat{\Omega}_n^{-1} \otimes (n^{-1}Z'Z)^{-1/2}\right) n^{-1/2} vec(Z'V) \\ &\to_d T^h \sim N(h_T, I_k). \end{aligned}$$

where $h_T = (a_0' h_{32}^{-1} a_0)^{1/2} (h_{31})^{1/2} h_1 h_2$. The convergence in distribution follows from LLN, CMT, CLT, Slutsky and homoskedastic assumptions. We now prove the asymptotic independence of $\widehat{S}_n, \widehat{T}_n$:

$$\begin{aligned} vec([\widehat{S}_n, \widehat{T}_n]) &= vec\left((n^{-1}Z'Z)^{1/2} n^{1/2} ||\pi_n|| \times \pi_n/||\pi_n|| a_0' \widehat{\Omega}_n^{-1/2} \widehat{J}\right) \\ &\quad + \left(\widehat{J}'\widehat{\Omega}_n^{-1/2} \otimes (n^{-1}Z'Z)^{-1/2}\right) n^{-1/2} vec(Z'V) \\ &\quad \to_d vec((h_{31})^{1/2} h_1 h_2 (h_{32})^{-1/2} h_J) + N(0, I_{2k}), \end{aligned}$$

where $h_J = [(h_{32})^{1/2} b_0/(b_0' h_{32} b_0)^{1/2}, (h_{32})^{-1/2} a_0/(a_0'(h_{32})^{-1} a_0)^{1/2}]$ .The convergence follows by the LLN, CMT, CLT, Slutsky and homoskedastic assumptions. Define the statistics:

$$Q^h_{(k-1)} = S^{h\prime} M^h_T S^h,\ Q^h_T = T^{h\prime} T^h \text{ and } LM1^h = S^{h\prime} T^h/(Q^h_T)^{1/2}.$$

The statistics $Q^h_{(k-1)}$ and $LM1^h$ are independent of Q^h_T. Now we can use the CMT to analyze the asymptotic distribution of the t-statistics. First we look for the $\underline{k}$-class estimators of β:

$$\begin{aligned}\widehat{\beta}(\widehat{\underline{k}}) = & \frac{\widehat{c}_{11}\widehat{c}_{21}\left(\widehat{Q}_{(k-1),n} + \widehat{LM1}_n^2\right) + (\widehat{c}_{12}\widehat{c}_{21} + \widehat{c}_{11}\widehat{c}_{22})\widehat{LM1}_n \widehat{Q}_{T,n}^{1/2}}{\widehat{c}_{21}^2\left(\widehat{Q}_{(k-1),n} + \widehat{LM1}_n^2\right) + 2\widehat{c}_{21}\widehat{c}_{22}\widehat{LM1}_n\widehat{Q}_{T,n}^{1/2} + \widehat{c}_{22}^2\widehat{Q}_{T,n} + n\left(1-\widehat{\underline{k}}\right)\widehat{\omega}_{22}} \\ & + \frac{\widehat{c}_{12}\widehat{c}_{22}\widehat{Q}_{T,n} + n\left(1-\widehat{\underline{k}}\right)\widehat{\omega}_{12}}{\widehat{c}_{21}^2\left(\widehat{Q}_{(k-1),n} + \widehat{LM1}_n^2\right) + 2\widehat{c}_{21}\widehat{c}_{22}\widehat{LM1}_n\widehat{Q}_{T,n}^{1/2} + \widehat{c}_{22}^2\widehat{Q}_{T,n} + n\left(1-\widehat{\underline{k}}\right)\widehat{\omega}_{22}}\end{aligned} \tag{4.46}$$

where $[\widehat{c}_{ij}] = \widehat{\Omega}_n^{1/2}\widehat{J}$ and the $\widehat{\underline{k}}$'s are given by:

$$\begin{array}{ll} \text{2SLS:} & \widehat{\underline{k}} = 1, \\ \text{LIML:} & \widehat{\underline{k}} = \widehat{\underline{k}}_{\,LIMLK} = 1 + n(\widehat{Q}_{S,n} - \widehat{LR}_n), \\ \text{Fuller:} & \widehat{\underline{k}} = \widehat{\underline{k}}_{LIMLK} - 1/n. \end{array}$$

For the 2SLS, LIML and Fuller estimators the term $n(1-\widehat{\underline{k}})$ converges in distribution to $\kappa^h(Q^h_{(k-1)}, LM^h_1, Q^h_T)$ where:

$$\begin{array}{ll} \text{2SLS:} & \kappa^h = 0, \\ \text{LIML:} & \kappa^h = \kappa^h_{LIML} = LR^h - Q^h_{(k-1)} - \left(LM^h_1\right)^2, \\ \text{Fuller:} & \kappa^h = \kappa^h_{LIML} + 1. \end{array}$$

where $LR^h = \frac{1}{2}\left(Q^h_{(k-1)} + \left(LM^h_1\right)^2 - Q^h_T + \sqrt{(Q_{(k-1)} + \left(LM^h_1\right)^2 - Q^h_T)^2 + 4\left(LM^h_1\right)^2\left(Q^h_T\right)}\right)$. We use the CMT in equation (4.46) to obtain:

$$\begin{aligned}\widehat{\beta}(\widehat{\underline{k}}) & \to_d \beta(\kappa^h) \\ & = \frac{c^h_{11}c^h_{21}\left(Q^h_{(k-1)} + \left(LM^h_1\right)^2\right) + (c^h_{12}c^h_{21} + c^h_{11}c^h_{22})LM^h_1\left(Q^h_T\right)^{1/2}}{\left(c^h_{21}\right)^2\left(Q^h_{(k-1)} + \left(LM^h_1\right)^2\right) + 2c^h_{21}c^h_{22}LM^h_1\left(Q^h_T\right)^{1/2} + \left(c^h_{22}\right)^2 Q^h_T + \kappa^h e_2' h_{32} e_2} \\ & + \frac{c^h_{12}c^h_{22}Q^h_T + \kappa^h e_1' h_{32} e_2}{\left(c^h_{21}\right)^2\left(Q^h_{(k-1)} + \left(LM^h_1\right)^2\right) + 2c^h_{21}c^h_{22}LM^h_1\left(Q^h_T\right)^{1/2} + \left(c^h_{22}\right)^2 Q^h_T + \kappa^h e_2' h_{32} e_2},\end{aligned}$$

where $[c^h_{ij}] = h_{32}^{1/2}h_J$. Now we use CMT again and obtain the asymptotic distribution of the t-statistics:

$$\begin{aligned}\widehat{t}(\widehat{\underline{k}}) & \to_d t(Q^h_{(k-1)}, LM^h_1, Q^h_T) \\ & = \frac{\left(\beta(\kappa^h) - \beta_0\right)\left(\sigma_u(\kappa^h)\right)^{-1}}{[\left(c^h_{21}\right)^2\left(Q^h_{(k-1)} + \left(LM^h_1\right)^2\right) + 2c^h_{21}c^h_{22}LM^h_1\left(Q^h_T\right)^{1/2} + \left(c^h_{22}\right)^2 Q^h_T + \kappa^h e_2' h_{32} e_2]^{-1/2}},\end{aligned}$$

where $\sigma^2_u(\kappa^h) = (1, -\beta(\kappa^h))h_{32}(1, -\beta(\kappa^h))'$. Next we see the asymptotic distribution of the critical value functions. The critical value function of the t-tests, $\kappa_{t(\underline{k}),\alpha}(q_T)$, is the $1-\alpha$ quantile of the conditional distribution:

$$t(q_T) = \frac{\beta\left(\underline{k}\right) - \beta_0}{\widehat{\sigma}_u\left(\underline{k}\right)\left[\widehat{c}_{21}^2\left(Q_{(k-1)} + LM_1^2\right) + 2\widehat{c}_{21}\widehat{c}_{22}LM_1 q_T^{1/2} + \widehat{c}_{22}^2 q_T + n(1-\underline{k})\widehat{w}_{22}\right]^{-1/2}}, \text{ where}$$

$$\beta\left(\underline{k}\right) = \frac{\widehat{c}_{11}\widehat{c}_{21}\left(Q_{(k-1)} + LM_1^2\right) + (\widehat{c}_{12}\widehat{c}_{21} + \widehat{c}_{11}\widehat{c}_{22})LM_1 q_T^{1/2} + \widehat{c}_{12}\widehat{c}_{22}q_T}{\widehat{c}_{21}^2\left(Q_{(k-1)} + LM_1^2\right) + 2\widehat{c}_{21}\widehat{c}_{22}LM_1 q_T^{1/2} + c_{22}^2 q_T}, \tag{4.47}$$

$\widehat{\sigma}_u(q_T) = (1, -\beta(\underline{k}))\widehat{\Omega}_n(1, -\beta(\underline{k}))'$ and $n(1-\underline{k})$ has distribution according to each estimator:

$$\begin{array}{ll} \text{2SLS:} & n(1-\underline{k}) = 0, \\ \text{LIMLK:} & n(1-\underline{k}) = \kappa_{\,LIMLK} = LR(q_T) - Q_{(k-1)} - LM_1^2 \\ \text{Fuller:} & n(1-\underline{k}) = \kappa_{LIMLK} - 1. \end{array}$$

where $LR(q_T) = \frac{1}{2}\left(Q_{(k-1)} + LM_1^2 - q_T + \sqrt{(Q_{(k-1)} + LM_1^2 - Q_T)^2 + 4LM_1^2 q_T}\right)$. Using the fact that the conditional distribution are continuous and the CMT we obtain:

$$\left(\widehat{t}(\underline{\widehat{k}})_n - \kappa_{t(\underline{k}),\alpha}(\widehat{Q}_{T,n})\right) \to_d \left(t(Q^h_{(k-1)}, LM^h_1, Q^h_T) - \kappa_{t(\underline{k}),\alpha}(Q^h_T)\right)$$

which is: $P_{\theta_n}(\widehat{t}(\underline{\widehat{k}})_n > \kappa_{t(\underline{k}),\alpha}(\widehat{Q}_{T,n})) \to P(t(Q^h_{(k-1)}, LM^h_1, Q^h_T) > \kappa_{t(\underline{k}),\alpha}(Q^h_T))$, but conditional on $Q^h_T = q_T$ the distribution of $t(Q^h_{(k-1)}, LM^h_1, Q^h_T)$ is the same as $t(q_T)$, hence, the conditional probability (on $Q^h_T = q_T$) of $t(Q^h_{(k-1)}, LM^h_1, Q^h_T) > \kappa_{t(\underline{k}),\alpha}(q_T)$ is exactly α. By the law of iterated expectations we have that the probability of $t(Q^h_{(k-1)}, LM^h_1, Q^h_T) > \kappa_{t(\underline{k}),\alpha}(q_T)$ is α. As we wanted to prove.

Case 2) $h_1 = \infty$. The asymptotic behavior of $\widehat{\Omega}_n$ and $\widehat{S}_n$ is the same as the case $h_1 < \infty$. For the statistics $\widehat{T}_n$ we need to normalize it by $n^{1/2}||\pi_n||$. Since $h_1 = \infty$, for n sufficiently large we have that $||\pi_n||$ is positive, so:

$$\begin{aligned} n^{-1/2}||\pi_n||^{-1}\widehat{T}_n &= (a_0'\widehat{\Omega}_n^{-1}a_0)^{1/2}(n^{-1}Z'Z)^{1/2}\pi_n/||\pi_n|| \\ &\quad + (a_0'\widehat{\Omega}_n^{-1}a_0)^{-1/2}(n^{-1}Z'Z)^{-1/2}(n^{-1/2}Z'V)n^{-1/2}||\pi_n||^{-1}\widehat{\Omega}_n^{-1}a_0 \\ &\to_p \lambda^h, \end{aligned}$$

with $\lambda^h = (a_0'h_{32}^{-1}a_0)^{1/2}{h_{31}}^{1/2}h_2$, this follows from LLN, CMT, CLT and using that $n^{-1/2}Z'V$ is bounded in probability and $n^{-1/2}||\pi_n||^{-1} = o(1)$. Consequently we can apply the CMT and Slutsky to obtain:

$$\begin{aligned} \widehat{Q}_{T,n}/n||\pi_n||^2 &= \widehat{T}_n'\widehat{T}_n/n||\pi_n||^2 \to_p \lambda^{h\prime}\lambda^h, \\ \widehat{Q}_{(k-1),n} &= \widehat{S}_n'(I - \widehat{T}_n(\widehat{T}_n'\widehat{T}_n)^{-1}\widehat{T}_n')\widehat{S}_n \to_d \chi_{(k-1)}, \\ \widehat{LM1}_n &= \left(\widehat{S}_n'\widehat{T}_n/n^{1/2}||\pi_n||\right)/(\left(\widehat{Q}_{T,n}/n||\pi_n||^2\right)^{1/2} \to_d N(0,1) \end{aligned}$$

Now we show that

$$n^{-1/2}||\pi_n||^{-1}n(1 - \underline{\widehat{k}}) \to_p 0 \tag{4.48}$$

for all k-class estimators used by MMV. For 2SL2S is trivial and for LIML and Fuller we just need to prove that (4.48) is true for LIML $\underline{\widehat{k}}$'s. Consider the analogous formulation of the statistic $\widehat{LR}_n$ used in equation 3.5 of MMV to derive:

$$\begin{aligned} &- n^{-1/2}||\pi_n||^{-1}n(\underline{\widehat{k}}_{LIML} - 1) \\ &= 2n^{-1/2}||\pi_n||^{-1}[\widehat{Q}_{S,n} + \widehat{Q}_{T,n} - \sqrt{(\widehat{Q}_{T,n} + \widehat{Q}_{S,n})^2 + 4(\widehat{Q}_{ST,n})^2}] \\ &= 2n^{-1/2}||\pi_n||^{-1}[(\widehat{Q}_{S,n} + \widehat{Q}_{T,n}) - |\widehat{Q}_{T,n} - \widehat{Q}_{S,n}|(1 + 4(\widehat{Q}_{ST,n})^2/(\widehat{Q}_{T,n} - \widehat{Q}_{S,n})^2)] + o_p(1) \\ &= o_p(1) \end{aligned}$$

where the second equality follows from a Taylor expansion of $f(x) = \sqrt{1+x}$ around $x = 0$ and the third equality from:

$$\frac{(\widehat{Q}_{ST,n})^2}{(\widehat{Q}_{T,n} - \widehat{Q}_{S,n})^2} = \frac{(n^{-1/2}||\pi_n||^{-1}\widehat{Q}_{ST,n})^2}{(n^{-1}||\pi_n||^{-2}\widehat{Q}_{T,n} - n^{-1}||\pi_n||^{-2}\widehat{Q}_{S,n})^2}\frac{1}{n||\pi_n||^2} \to_p 0$$

Hence, $n^{-1/2}||\pi_n||^{-1}n(1 - \underline{\widehat{k}}_{LIML}) \to_p 0$ and we obtain (4.48). Using this result obtain that the k-class estimator $\widehat{\beta}(\underline{\widehat{k}})$ converges in probability to β_0. Next, we can obtain the asymptotic distribution of the t-statistics:

$$\begin{aligned} \widehat{t}(\underline{\widehat{k}})_n &= \frac{[\widehat{c}_{21}\left(\widehat{Q}_{(k-1),n} + \widehat{LM}^2_{1,n}\right) + \widehat{c}_{22}\widehat{LM}_{1,n}\widehat{Q}^{1/2}_{T,n}]}{[\widehat{c}^2_{21}\left(\widehat{Q}_{(k-1),n} + \widehat{LM}^2_{1,n}\right) + 2\widehat{c}_{21}\widehat{c}_{22}\widehat{LM}_{1,n}\widehat{Q}^{1/2}_{T,n} + \widehat{c}^2_{22}\widehat{Q}^{1/2}_{T,n} + n\left(1 - \underline{\widehat{k}}\right)\widehat{\omega}_{22}]^{1/2}} \cdot \frac{\widehat{\sigma}_0}{\widehat{\sigma}_u(\underline{\widehat{k}})} \\ &\quad + \frac{e_2'\widehat{\Omega}_n b_0 n(1 - \underline{\widehat{k}})}{\widehat{\sigma}_u(\underline{\widehat{k}})[\widehat{c}^2_{21}\left(\widehat{Q}_{(k-1),n} + \widehat{LM}^2_{1,n}\right) + 2\widehat{c}_{21}\widehat{c}_{22}\widehat{LM}_{1,n}\widehat{Q}^{1/2}_{T,n} + \widehat{c}^2_{22}\widehat{Q}^{1/2}_{T,n} + n\left(1 - \underline{\widehat{k}}\right)\widehat{\omega}_{22}]^{1/2}} \\ &\quad \to_d LM^h_1 \sim N(0,1). \end{aligned}$$

Finally, we analyze the critical value function. Because $q_T \to_p \infty$ as $n \to \infty$, the conditional distributions of t-statistics, equation (4.47), converge to a standard normal distribution; see Lemma 1 of MMV. Hence, the $1-\alpha$ quantile of the conditional distribution of $t(q_T)$ converges in probability to the $1-\alpha$ quantile of the standard normal distribution ($z_{1-\alpha}$). Therefore, from the definition of convergence in distribution,

$$P_{\theta_n}(\widehat{t}(\widehat{\underline{k}})_n > \kappa_{t(\underline{k}),\alpha}(\widehat{Q}_{T,n})) \to \Phi(Z > z_{1-\alpha}) = \alpha$$

where Φ is the cumulative distribution of the normal distribution. The proof is complete. ■

4.2 Proofs of Chapter 2

Proof. [Proof of (11)] Let $\{\Theta_e\}_{e\in E}$ partition of Θ_R and ϕ^* General Maxmin in Θ_e in the assumptions and define $\beta_e^* := \beta^*(\theta)$ for all $\theta \in \Theta_e$. Then for all $\phi \in \Gamma(g,\gamma)$:

$$\sup_{\theta\in\Theta_R}\left(\beta^*(\theta) - \int_\Omega \phi f_\theta d\mu\right) = \sup_{e\in E}\left(\beta_e^* - \inf_{\theta\in\Theta_e}\int_\Omega \phi f_\theta d\mu\right) \tag{4.49}$$

$$\geq \sup_{e\in E}\left(\beta_e^* - \inf_{\theta\in\Theta_e}\int_\Omega \phi^* f_\theta d\mu\right) \tag{4.50}$$

$$= \sup_{\theta\in\Theta_R}\left(\beta^*(\theta) - \int_\Omega \phi^* f_\theta d\mu\right). \tag{4.51}$$

Hence ϕ^* is the General Minimax Regret test. ■

Proof. [Proof of (12)] **a)** Note that the set of tests functions $\Gamma(g,\gamma)$ is a finite intersection of weakly-star closed subsets of the closed unit ball $B_{L_\mu^\infty} := \{\phi \in L_\mu^\infty : ||\phi||_\infty \leq 1\}$, that is weakly-star compact by Banach-Alaoglu-Bourbanki Theorem (Theorem 3.16 in Brezis (2011)). Then, $\Gamma(g,\gamma)$ is a weakly-star compact set. Since

$$\{(\phi,a) \in \Gamma(g,\gamma)\times\mathbb{R} : -\inf_{\theta\in\Theta_R}\int_\Omega \phi f_\theta d\mu \leq a\}$$

$$= \bigcap_{\theta\in\Theta_R}\{(\phi,a)\in\Gamma(g,\gamma)\times\mathbb{R} : -\int_\Omega \phi f_\theta d\mu \leq a\}$$

is an arbitrary intersection of weakly-star closed sets, the functional

$$\Gamma(g,\gamma) \ni \phi \longmapsto \inf_{\theta\in\Theta_R}\int_\Omega \phi f_\theta d\mu,$$

is weakly-star upper semicontinuous. Hence, the existence of the GM test follows from the Weierstrass Theorem.

b) Here we follow closelly Krafft and Witting (1967) in stating the GMP as an Infinite Linear Programming and deriving the Fenchel-Rockafellar Dual Problem. Note that the GM problem (2.7) can be written as

$$\begin{aligned} &\sup_{(\phi,r)\,\in\, L_{\mu,+}^\infty\times\mathbb{R}_+,\ \text{s.t}} \quad Id(r) + \underline{0}(\phi) \\ &-r + \int_\Omega \phi f_\theta d\mu \geq 0,\ \forall\theta\in\Theta_R \\ &-\gamma_{k,1}(\theta) + \int_\Omega \phi g_\theta^k d\mu \geq 0,\ \forall\theta\in\Theta_{k,S},\ \text{for } k=1,...,l \\ &\gamma_{k,2}(\theta) - \int_\Omega \phi g_\theta^k d\mu \geq 0,\ \forall\theta\in\Theta_{k,S},\ \text{for } k=1,...,l \\ &\underline{1}(w) - \phi(w) \geq 0, \forall w\in\Omega \end{aligned} \tag{4.52}$$

where $Id(x) = x$, $\underline{0}(x) = 0$, $\underline{1}(x) = 1$ functionals and $L_{\mu,+}^\infty := \{\phi \in L_\mu^\infty : \phi \geq 0\ \mu - a.e.\}$. In order to derive the Dual Problem we use the following notation of functionals:

$$\langle x^*, x\rangle := x^*(x),$$

where $x^* \in B^*$, with B^* Dual Space of the Banach Space B. Note that the Primal Problem (4.52), denoted by $\boldsymbol{P}$, can be written as:

$$\sup_{x\in C,\ b-A(x)\in K} \langle c, x\rangle, \tag{4.53}$$

where $X := L_\mu^\infty \times \mathbb{R}$ and $Y := Y_1\times Y_2\times Y_3$, $Y_1 := C(\Theta_R)$, $Y_2 := \left(\times_{k=1}^l C(\Theta_{k,S})\right)^2$, $Y_3 := L_\mu^\infty$, where $C(\mathcal{X})$ is the Banach Space of real continuous functions defined on the compact metric space $\mathcal{X}$; the continuous linear

functional $c := (\underline{0}, \underline{1}) \in X^*$; the vector $b := (\underline{0}, (-\gamma_{k,1})_{k=1}^l, (\gamma_{k,2})_{k=1}^l, \underline{1}) \in Y$ and the bounded linear mapping $A : X \longmapsto Y$,

$$A(\phi, r) = \left(r - \int_\Omega \phi f_\theta d\mu, \left(-\int_\Omega \phi g_\theta^k d\mu \right)_{k=1}^l, \left(\int_\Omega \phi g_\theta^k d\mu \right)_{k=1}^l, \phi \right).$$

Indeed A is well defined since for every $\phi \in L_\mu^\infty$: i) $\theta \mapsto \int_\Omega \phi f_\theta d\mu \in C(\Theta_R)$ because $|\int_\Omega \phi f_{\theta_n} d\mu - \int_\Omega \phi f_\theta d\mu| \leq ||\phi||_{L_\mu^\infty} \times \int_\Omega |f_{\theta_n} - f_\theta| d\mu \to 0$ for every convergent sequence $\theta_n \to \theta$ by Scheffé Theorem (Lemma 2.1 in Bauer (2001)); ii) $\theta \mapsto \int_\Omega \phi g_\theta^k d\mu \in C(\Theta_S)$ for all $k = 1, ..., l$ by *Assumption 4* and Scheffé Theorem (Lemma 2.1 in Bauer (2001)). The constraints of the GMP are posed in terms of the closed convex cones: $C := L_{\mu,+}^\infty \times \mathbb{R}_+$ and $K := C(\Theta_R)_+ \times (\times_{k=1}^l C(\Theta_{k,S})_+)^2 \times L_{\mu,+}^\infty$, where $C(\mathcal{X})_+ := \{f \in C(\mathcal{X}) : f \geq 0\}$.

Stated the GMP as an Infinite Linear Programming Problem, we can use the duality approach to solve this problem. Consider the negative of the problem (4.52), $(-\boldsymbol{P})$:

$$\inf_{x \in C,\ b - A(x) \in K} \langle c', x \rangle, \tag{4.54}$$

where $c' := -c$. Associated to this problem, we consider the pertubed function: $\Phi : X \times Y \to \mathbb{R}$, $\Phi(x, y) = \langle c', x \rangle + I_C(x) + I_K(b - A(x) - y)$ where $I_Q(x) = 0$ if $x \in Q$ and $+\infty$ otherwise; and the pertubed problem:

$$h(y) := \inf_{x \in X} \Phi(x, y). \tag{4.55}$$

Clearly, $h(0) = v(-\boldsymbol{P})$, where $v(\cdot)$ is the optimal value of the problem $(\cdot)$. The Fenchel-Rockafellar Dual Problem of $(-\boldsymbol{P})$, $(-\boldsymbol{P}^*)$, is defined as

$$\sup_{y^* \in Y^*} -\Phi^*(0, y^*), \tag{4.56}$$

where $\Phi^* : X^* \times Y^* \to \overline{\mathbb{R}}$, $\Phi^*(x^*, y^*) := \sup_{(x,y) \in X \times Y} [\langle (x, y), (x^*, y^*) \rangle - \Phi(x, y)]$ is the Legendre transformation of Φ. Note that

$$\begin{aligned} \Phi^*(0, y^*) &= \sup_{x \in C,\ b - A(x) - y \in K} (\langle y, y^* \rangle - \langle c', x \rangle) \\ &= \sup_{x \in C} \left[\sup_{y \in K} (\langle b, y^* \rangle - \langle y, y^* \rangle - \langle A^* y^* + c', x \rangle) \right] \\ &= \langle b, y^* \rangle + I_{K^+}(y^*) + I_{C^+}(A^* y^* + c'), \end{aligned} \tag{4.57}$$

where $A^* : Y^* \to X^*$ is the adjoint operator of A, $K^+ := \{y^* \in Y^* : \langle y, y^* \rangle \geq 0$ for all $y \in K\}$ and $C^+ := \{x^* \in Y^* : \langle x, x^* \rangle \geq 0$ for all $x \in C\}$. Hence, the Dual Problem $(-\boldsymbol{P}^*)$ is

$$\sup_{y^* \in K^+,\ A^* y^* + c' \in C^+} -\langle b, y^* \rangle. \tag{4.58}$$

Since we considered the minus primal problem $(-\boldsymbol{P}^*)$ and $c' = -c$, the Dual Problem $(\boldsymbol{P}^*)$ of the Primal Problem $(\boldsymbol{P})$ is:

$$\inf_{y^* \in K^+,\ A^* y^* - c \in C^+} \langle b, y^* \rangle, \tag{4.59}$$

that is denoted as $(\boldsymbol{P}^*)$. From the Riesz Representation Theorem, the Dual Space of the space of continuous functions defined on the compact metric space $\mathcal{X}$, $C(\mathcal{X})$, is the space of all regular Borel signed measures $\mathcal{M}^s(\mathcal{X})$ (see Theorem C.18 in Conway (1990)). The Dual Space of the space L_μ^∞ is $ba(\Omega, \mathcal{F}, \mu)$, space of all bounded, finitely additive set-functions wich are absolute continuous with respect to μ (see Chapter IV, Section 9 in Yosida (1995)). Hence, the constraints in the Dual Problem $(\boldsymbol{P}^*)$ are

$$y^* := (\lambda, (\eta_{k,1})_{k=1}^l, (\eta_{k,2})_{k=1}^l, \xi) \in \mathcal{M}(\Theta_R) \times \left(\times_{k=1}^l \mathcal{M}(\Theta_{k,S}) \right)^2 \times ba(\Omega, \mathcal{F}, \mu)_+, \quad \text{and} \tag{4.60}$$

$$r(\lambda(\Theta_R) - 1) - \int_{\Theta_R} \int_\Omega \phi f_\theta d\mu d\lambda + \sum_{k=1}^l \int_{\Theta_{k,S}} \int_\Omega \phi g_\theta^k d\mu d\left(\eta_{k,2} - \eta_{k,1} \right) + \int_\Omega \phi \frac{d\xi}{d\mu} d\mu \geq 0 \tag{4.61}$$

for all $(\phi, r) \in C$ and the objective function is

$$\langle b, y^* \rangle = \sum_{k=1}^{l} \left(\int_{\Theta_{k,S}} \gamma_{k,2}(\theta) d\eta_{k,2} - \int_{\Theta_{k,S}} \gamma_{k,1}(\theta) d\eta_{k,1} \right) + \int_{\Omega} \frac{d\xi}{d\mu} d\mu \tag{4.62}$$

Using Fubini-Tonelli Theorem in (4.61) and $r = 0$ we obtain

$$\frac{d\xi}{d\mu} \geq \int_{\Theta_R} f_\theta d\lambda - \sum_{k=1}^{l} \int_{\Theta_S} g_\theta d(\eta_{k,2} - \eta_{k,1}),\ \mu - a.e., \tag{4.63}$$

combining with the fact that the objective function is nondecreasing in $\frac{d\xi}{d\mu} \geq 0$ we obtain the optimal

$$\frac{d\xi^*}{d\mu} = \left(\int_{\Theta_R} f_\theta d\lambda - \sum_{k=1}^{l} \int_{\Theta_{k,S}} g_\theta^k d(\eta_{k,2} - \eta_{k,1}) \right)^+ \ \mu - a.e. \tag{4.64}$$

Substitute $\phi = 0$ in (4.61) and obtain that $\lambda(\Theta_R) \geq 1$. Furthermore, the objective function is nondecreasing in λ, so $\lambda(\Theta_R) = 1$. Substituting (4.64) in the Dual Problem (4.59) we obtain (2.13):

$$\begin{aligned} \inf_{(\lambda, (\eta_{k,1}, \eta_{k,2})_{k=1}^l) \in (\Lambda, \Xi)} & \sum_{k=1}^{l} \left(\int_{\Theta_{k,S}} \gamma_{k,2}(\theta) d\eta_{k,2} - \int_{\Theta_{k,S}} \gamma_{k,1}(\theta) d\eta_{k,1} \right) \\ & + \int_{\Omega} \left(\int_{\Theta_R} f_\theta d\lambda - \sum_{k=1}^{l} \int_{\Theta_{k,S}} g_\theta^k d(\eta_{k,2} - \eta_{k,1}) \right)^+ d\mu. \end{aligned}$$

To prove the Strong Duality and No Gap properties we just need to prove the Generalized Slater Condition

$$\text{there exists } x_0 \in C \text{ such } b - A(x_0) \in \text{interior}(K), \tag{4.65}$$

where the interior is with respect to the cartesian norm: $||.|| := ||.||_{Y_1} + ||.||_{Y_2} + ||.||_{Y_3}$. Indeed, if equation (4.65) is satisfied, exist $\epsilon > 0$ such $b - A(x_0) - y \in K$ for all $y \in B_Y(0, \epsilon) := \{y \in Y : ||y|| \leq \epsilon\}$. Hence, $y \mapsto \Phi(x_0, y) = \langle c', x_0 \rangle$ is finite for all $y \in B_Y(0, \epsilon)$ and

$$h(y) = \inf_{x \in X} \Phi(x, y) \leq \Phi(x_0, y) = \langle c', x_0 \rangle \text{ for all } y \in B_Y(0, \epsilon), \tag{4.66}$$

Since the convex function h is bounded above in a neighborhood of the origin, the subdifferential of h at 0 is nonempty

$$\partial h(0) := \{y^* \in Y^* : h(0) - h(y) \leq \langle y^*, 0 - y \rangle \text{ for all } y \in Y\} \neq \emptyset, \tag{4.67}$$

(Propositions 2.5 and 5.2 of Chapter I in Ekeland and Teman (1999)). Its well known in convex analysis that $\partial h(0) \neq \emptyset$ is sufficient for

$$h(0) = h^{**}(0) \text{ and } \partial h(0) = \partial h^{**}(0) \neq \emptyset, \tag{4.68}$$

(Equations 5.3 and 5.4 of Chapter I in Ekeland and Teman (1999)), where $h^{**}(0) = v(-\boldsymbol{P}^*)$ and $\partial h^{**}(0)$ is the set of solutions of the Dual Problem $(\boldsymbol{P}^*)$ (Lemma 2.3 and 2.4 Chapter III in Ekeland and Teman (1999)). Hence, we have stablish

$$\text{No Gap property} \ : \ v(\mathcal{P}^*) = h^{**}(0) = h(0) = v(\mathcal{P}) \text{ and} \tag{4.69}$$

$$\text{Strong Duality} \ : \ \text{solutions}(\mathcal{P}^*) = \partial h^{**}(0) \neq \varnothing. \tag{4.70}$$

To prove the Generalized Slater Condition (4.65) let ϕ_0 be the test in *Assumption 1*, and note that $x_0 = (\phi_0, 0) \in C$ is such

$$\begin{aligned} & b - A(x_0) \\ = & \left(\int_{\Omega} \phi_0 f_\theta d\mu, \left(\int_{\Omega} \phi_0 g_\theta d\mu - \gamma_{k,1}(\theta)_{k=1}^l \right)_{k=1}^l, \left(\gamma_{k,2}(\theta) - \int_{\Omega} \phi_0 g_\theta d\mu_{k=1}^l \right)_{k=1}^l, 1 - \phi_0 \right) \\ \in & \ \text{interior}(K), \end{aligned} \tag{4.71}$$

since
$$\int_\Omega \phi_0 f_\theta d\mu > 0;\ \gamma_{k,1}(\theta) < \int_\Omega \phi_0 g_\theta d\mu < \gamma_{k,2}(\theta) \text{ for all } \theta \in \Theta_{k,S},\ k = 1, ..., l; \text{ and } \phi_0 < 1.$$
Now we prove the representation of the GM test. For any $\phi \in \Gamma(g, \gamma)$ and $(\lambda, (\eta_{k,1}, \eta_{k,2})_{k=1}^l) \in (\Lambda, \Xi)$ note that

$$\begin{aligned}
&\inf_{\theta\in\Theta_R} \int_\Omega \phi f_\theta d\mu \\
&\leq \int_{\Theta_R} \int_\Omega \phi f_\theta d\mu d\lambda \\
&\leq \int_{\Theta_R} \int_\Omega \phi f_\theta d\mu d\lambda + \sum_{k=1}^l \int_{\Theta_{k,S}} \left(\int_\Omega \phi g_\theta^k d\mu - \gamma_{k,1}(\theta) \right) d\eta_{k,1} \\
&\leq \int_{\Theta_R} \int_\Omega \phi f_\theta d\mu d\lambda + \sum_{k=1}^l \int_{\Theta_{k,S}} \left(\int_\Omega \phi g_\theta^k d\mu - \gamma_{k,1}(\theta) \right) d\eta_{k,1} + \sum_{k=1}^l \int_{\Theta_{k,S}} \left(\gamma_{k,2}(\theta) - \int_\Omega \phi g_\theta^k d\mu \right) d\eta_{k,2} \\
&\leq \sum_{k=1}^l \left(\int_{\Theta_{k,S}} \gamma_{k,2}(\theta) d\eta_{k,2} - \int_{\Theta_{k,S}} \gamma_{k,1}(\theta) d\eta_{k,1} \right) + \int_\Omega \left(\int_{\Theta_R} f_\theta d\lambda - \sum_{k=1}^l \int_{\Theta_{k,S}} g_\theta^k d(\eta_{k,2} - \eta_{k,1}) \right)^+ d\mu. \quad (4.72)
\end{aligned}$$

The No Gap property (4.69) implies that all inequalities in (4.72) holds as equalities at the solutions $\phi^* \in \Gamma(g, \gamma)$ and $(\lambda^*, (\eta_{k,1}^*, \eta_{k,1}^*)_{k=1}^l) \in (\Lambda, \Xi)$. Hence, the proof is complete. ■

Proof. [Proof of (13)] Set $g_\theta^k = f_\theta$ for all $\theta \in \Theta_R = \Theta_K$, $\gamma_{k,1} = 0$, for all $k = 1, ..., l$, $\gamma_{1,2} = \alpha$ and $\gamma_{k,2} = 0$ for all $k = 2, ..., l$, $\Theta_S = \Theta_H$ and note that $\phi_0 = \alpha \in \Gamma(g, \gamma) = \Phi_{\alpha,H}$. Then the existence follows from Theorem 1 a). For Theorem 1 letter b) we just need to prove that *Assumption 1* and *4* are valid. *Assumption 1* is true for $\widetilde{\phi} = \phi_0/2 \in \Phi_{\alpha,H}$, since $\phi_0 \in (0, 1)$ and $\int_\Omega \phi_0 f_\theta d\mu \in (0, \alpha)$ for all $\theta \in \Theta_H$. *Assumption 3* is valid from A.3') and Scheffé Theorem (Lemma 2.1 in Bauer (2001)). ■

Proof. [Proof of (14)] **a)** From Theorem (12) a), exist $\phi_\theta \in \Gamma(g, \gamma)$ for all $\theta \in \Theta_R$ such
$$\int_\Omega \phi_\theta f_\theta d\mu = \sup_{\phi\in\Gamma(g,\gamma)} \int_\Omega \phi f_\theta d\mu. \quad (4.73)$$
Then, the power envelope
$$\Theta_R \ni \theta \mapsto \sup_{\phi\in\Gamma(g,\gamma)} \int_\Omega \phi f_\theta d\mu, \quad (4.74)$$
is well defined. Since
$$\begin{aligned}
&\{(\phi, a) \in \Gamma(g, \gamma, \delta) \times \mathbb{R} : \sup_{\theta\in\Theta_R} \left(\beta^*(\theta) - \int_\Omega \phi f_\theta d\mu \right) \leq a\} \\
&= \bigcap_{\theta\in\Theta_R} \{(\phi, a) \in \Gamma(g, \gamma, \delta) \times \mathbb{R} : \beta^*(\theta) - \int_\Omega \phi f_\theta d\mu \leq a\} \quad (4.75)
\end{aligned}$$
is a weakly-star closed set, the functional
$$\Gamma(g, \gamma) \ni \phi \longmapsto \sup_{\theta\in\Theta_R} \left(\beta^*(\theta) - \int_\Omega \phi f_\theta d\mu \right) \quad (4.76)$$
is weakly-star lower semicontinuous function over a weakly-star compact set $\Gamma(g, \gamma)$. Then the existence of the GMR test follows from the Weierstrass Theorem.

b) Note that
$$\begin{aligned}
&\inf_{\phi\in\Gamma(g,\gamma)} \sup_{\theta\in\Theta_R} \left(\beta^*(\theta) - \int_\Omega \phi f_\theta d\mu \right) \\
&= \inf_{\phi\in\Gamma(g,\gamma)} \sup_{\lambda\in\mathcal{M}_1(\Theta_R)} \left(\int_{\Theta_R} \beta^*(\theta) d\lambda - \int_\Omega \phi \left(\int_{\Theta_R} f_\theta d\lambda \right) d\mu \right). \quad (4.77)
\end{aligned}$$

Indeed, for every $\lambda \in \mathcal{M}_1(\Theta_R)$

$$\sup_{\theta\in\Theta_R}\left(\beta^*(\theta)-\int_\Omega \phi f_\theta d\mu\right) \geq \int_{\Theta_R}\beta^*(\theta)d\lambda - \int_\Omega \phi\left(\int_{\Theta_R} f_\theta d\lambda\right)d\mu, \tag{4.78}$$

and

$$\begin{aligned} & \sup_{\theta\in\Theta_R}\left(\beta^*(\theta)-\int_\Omega \phi f_\theta d\mu\right) \\ = & \sup_{\lambda\in\left\{\lambda\in\mathcal{M}_1(\Theta_R):\lambda=\chi_{\{\theta\}},\theta\in\Theta_R\right\}}\left(\int_{\Theta_R}\beta^*(\theta)d\lambda-\int_{\Theta_R}\int_\Omega \phi f_\theta d\mu d\lambda\right) \\ \leq & \sup_{\lambda\in\mathcal{M}_1(\Theta_R)}\left(\int_{\Theta_R}\beta^*(\theta)d\lambda-\int_\Omega\phi\left(\int_{\Theta_R} f_\theta d\lambda\right)d\mu\right), \end{aligned}$$

where $\chi_{\{\theta\}}(\cdot)$ is the Dirac Measure on $\theta \in \Theta_R$. Now we apply the Sion Minimax Theorem (Theorem 3.4 in Sion (1958)) in the right side of (4.77). For these we need to prove the Sion Minimax Assumptions: **b.i)** $\Gamma(g,\gamma)$ is a weakly-star compact and convex subset of L^∞_μ, **b.ii)** $\mathcal{M}_1(\Theta_R)$ is a weakly-star compact and convex subset of $\mathcal{M}(\Theta_R)$;[1] **b.iii)** the function

$$\Gamma(g,\gamma)\ni\phi\mapsto\int_{\Theta_R}\beta^*(\theta)d\lambda-\int_\Omega\phi\left(\int_{\Theta_R} f_\theta d\lambda\right)d\mu \tag{4.79}$$

is convex and weakly-star lower semicontinuous; **b.iv)** the function

$$\mathcal{M}_1(\Theta_R)\ni\lambda\mapsto\int_{\Theta_R}\beta^*(\theta)d\lambda-\int_\Omega\phi\left(\int_{\Theta_R} f_\theta d\lambda\right)d\mu \tag{4.80}$$

is concave and weakly-star upper semicontinuous. For the proof of **b.i)** and **b.iii)** see Theorem (12) proof's. **b.ii)** convexity of $\mathcal{M}_1(\Theta_R)$ follows trivially; for compactness note that $\mathcal{M}_1(\Theta_R)$ is a weakly-star closed set of the unit ball in $\mathcal{M}(\Theta_R)$, that is weakly-star compact by the Banach-Alaoglu-Bourbanki Theorem (Theorem 3.16 in Brezis (2011)). **b.iv)** Concave follows from linearity. For weakly-star lower semicontinuous, we prove that its weakly-star continuous. Fix $(\lambda^n)_{n=1}^\infty \in \mathcal{M}_1(\Theta_R)$ such λ^n weakly-star conveges to λ, denoted as $\lambda^n \rightharpoonup^* \lambda$. First, note that the power envelope is continuous function of the parameter, $\theta\mapsto\beta^*(\theta)\in C(\Theta_R)$, since $\phi\mapsto\int_\Omega \phi f_\theta d\mu$ is continuous in the weak-star topology for all $\forall\theta\in\Theta_R$ and $\Gamma(g,\gamma)$ is weakly-star compact. Second, from the definition of weak-star convergence of the probability measures $(\lambda^n)_{n=1}^\infty$ we conclude

$$\int_{\Theta_R}\beta^*(\theta)d\lambda^n\to\int_{\Theta_R}\beta^*(\theta)d\lambda.$$

Using Tonelli-Fubini and continuity of the power function we obtain

$$\int_\Omega\phi\left(\int_{\Theta_R} f_\theta d\lambda^n\right)d\mu\to\int_\Omega\phi\left(\int_{\Theta_R} f_\theta d\lambda\right)d\mu$$

for every $\phi\in\Gamma(g,\gamma)$. Hence, the function

$$\lambda\mapsto\int_{\Theta_R}\beta^*(\theta)d\lambda-\int_\Omega\phi\left(\int_{\Theta_R} f_\theta d\lambda\right)d\mu$$

is weakly-star continuous. Then we can apply the Sion Minimax Theorem in (4.77):

$$\begin{aligned} & \inf_{\phi\in\Gamma(g,\gamma)}\sup_{\theta\in\Theta_R}\left(\beta^*(\theta)-\int_\Omega\phi f_\theta d\mu\right) \\ = & \sup_{\lambda\in\mathcal{M}_1(\Theta_R)}\inf_{\phi\in\Gamma(g,\gamma)}\int_{\Theta_R}\beta^*(\theta)d\lambda-\int_\Omega\phi\left(\int_{\Theta_R} f_\theta d\lambda\right)d\mu \end{aligned} \tag{4.81}$$

[1] For Sion Minimax Theorem is not necessarily to $\mathcal{M}_1(\Theta_R)$ be weakly-star compact (see Corollary 3.3 in Sion (1958)), here we use to guaratee the existence for the solution of $\sup_{\lambda\in\mathcal{M}_1(\Theta_R)}\left(\int_{\Theta_R}U(\theta)d\lambda-\int_\Omega\phi\left(\int_{\Theta_R} f_\theta d\lambda\right)d\mu\right)$.

Note that the infimum problem in (4.81) is a special case of GMP, then the Strong Duality and No Gap Properties in Theorem (12) give

$$
\begin{aligned}
&\inf_{\phi\in\Gamma(g,\gamma)} \sup_{\theta\in\Theta_R} \left(\beta^*(\theta) - \int_\Omega \phi f_\theta d\mu\right) \\
= &\sup_{(\lambda,(\eta_{k,1},\eta_{k,2})_{k=1}^l)\in(\Lambda,\Xi)} \int_{\Theta_R} \beta^*(\theta)d\lambda + \sum_{k=1}^l \left(\int_{\Theta_{k,S}} \gamma_{k,1}(\theta)d\eta_{k,1} - \int_{\Theta_{k,S}} \gamma_{k,2}(\theta)d\eta_{k,2}\right) \\
&\qquad - \int_\Omega \left(\int_{\Theta_R} f_\theta d\lambda - \sum_{k=1}^l \int_{\Theta_{k,S}} g_\theta^k d(\eta_{k,2}-\eta_{k,1})\right)^+ d\mu. \qquad (4.82)
\end{aligned}
$$

Finally we characterize the GMR test $\phi^\dagger \in \Gamma(g,\gamma)$. Let $(\lambda,(\eta_{k,1},\eta_{k,2})_{k=1}^l) \in (\Lambda,\Xi)$ and $\phi\in\Gamma(g,\gamma)$, using the Tonelli-Fubini Theorem we note that

$$
\begin{aligned}
&\sup_{\theta\in\Theta_R} \left(\beta^*(\theta) - \int_\Omega \phi f_\theta d\mu\right) \\
\geq &\int_{\Theta_R} \beta^*(\theta)d\lambda - \int_\Omega \phi \left(\int_{\Theta_R} f_\theta d\lambda\right) d\mu \\
\geq &\int_{\Theta_R} \beta^*(\theta)d\lambda - \int_\Omega \phi \left(\int_{\Theta_R} f_\theta d\lambda\right) d\mu - \sum_{k=1}^l \int_{\Theta_{k,S}} (\int_\Omega \phi g_\theta^k d\mu - \gamma_{k,1}(\theta))d\eta_{k,1} \\
\geq &\int_{\Theta_R} \beta^*(\theta)d\lambda + \sum_{k=1}^l \left(\int_{\Theta_S} \gamma_{k,1}(\theta)d\eta_{k,1} - \int_{\Theta_S} \gamma_{k,2}(\theta)d\eta_{k,2}\right) \\
&\quad - \int_\Omega \phi \left(\int_{\Theta_R} f_\theta d\lambda - \sum_{k=1}^l \int_{\Theta_S} g_\theta d(\eta_{k,2}-\eta_{k,1})\right) d\mu \\
\geq &\int_{\Theta_R} \beta^*(\theta)d\lambda + \sum_{k=1}^l \left(\int_{\Theta_S} \gamma_{k,1}(\theta)d\eta_{k,1} - \int_{\Theta_S} \gamma_{k,2}(\theta)d\eta_{k,2}\right) \\
&\quad - \int_\Omega \left(\int_{\Theta_R} f_\theta d\lambda - \sum_{k=1}^l \int_{\Theta_S} g_\theta d(\eta_{k,2}-\eta_{k,1})\right)^+ d\mu. \qquad (4.83)
\end{aligned}
$$

Since exist $\phi^\dagger \in \Gamma(g,\gamma)$ and $\left(\lambda^\dagger,(\eta_{k,1}^\dagger,\eta_{k,2}^\dagger)_{k=1}^l\right) \in (\Lambda,\Xi)$ that solve the right and left side of (4.82) and the equality (Strong Duality) holds, then all inequalities in (4.83) hold as equalities. Concluding that the GMR test $\phi^\dagger\in\Gamma(g,\gamma)$ satisfies equations (2.18), (2.19), (2.20), (2.21), (2.22). ■

Proof. [Proof of (15)] The proof of Corollary 15 follows from the same arguments as in the proof of (13), thus is ommited. ■

Proof. [Proof of (16)] For each $\phi\in\Gamma(g,\gamma)$ we can use Assumption I.2 and define the sequence of tests $\phi_n := \int_G (\phi\circ u_\tau)\, d\mu_{n,H}(\tau)$, where $\mu_{n,H}(\cdot) = \mu_H(\cdot\cap K_n)/\mu_H(K_n)$ and K_n compacts in assumption I.2. First, we prove that exist a subsequence $(\phi_{n_j})_j$ and an $\mathcal{P}$-almost invariant test $\phi_I\in\Gamma(g,\gamma)$ such ϕ_{n_j} weakly-star converges to ϕ_I. Fix a $\theta\in\Theta_{k,S}$. Using Assumption I.1, Fubini-Tonelli and invariance of the densities we obtain:

$$
\begin{aligned}
\int_\Omega \phi_n g_\theta^k d\mu &= \int_G \int_\Omega \phi f_{v_\tau(\theta)} m_k d\mu d\mu_{n,H} \\
&\leq \sup_{\theta'\in Orb(\theta)} \int_\Omega \phi f_{v_\tau(\theta)} m_k d\mu, \qquad (4.84)
\end{aligned}
$$

where $Orb(\theta) := \{v_\tau(\theta) : \tau\in G\} \subseteq \Theta_{k,S}$ orbit of θ. Since:

$$\int_\Omega \phi f_{v_\tau(\theta)} m_k d\mu \leq \gamma_{2,k}(v_\tau(\theta)) = \gamma_{2,k}(\theta),$$

equation (4.84) is bounded by $\gamma_{2,k}(\theta)$. The same arguments proves that:

$$\gamma_{1,k}(\theta) \leq \int_{\Omega} \phi_n g_\theta^k d\mu.$$

Hence $(\phi_n)_n \subseteq \Gamma(g,\gamma)$ and by Banach-Alaoglu-Bourbanki Theorem (Theorem 3.16 in Brezis (2011)), exist a subsequence $\left(\phi_{n_j}\right)_j$ and a limit $\phi_I \in \Gamma(g,\gamma)$ such:

$$\int_{\Omega} \phi_{n_j} h d\mu \to \int_{\Omega} \phi_I h d\mu,$$

for any $h \in L^1_\mu$. Furthermore, the Hunt-Stein Theorem proves that the limit test ϕ_I is $\mathcal{P}$-almost invariant. Now we prove the first statement of the Corollary. Using Fubini-Tonelli and invariance of the densities again we obtain:

$$\begin{aligned}
\inf_{\theta \in \Theta_R} \int_{\Omega} \phi_I f_\theta d\mu &= \inf_{\theta \in \Theta_R} \lim_{n_j \to \infty} \int_G \int_{\Omega} \phi f_{v_\tau(\theta)} d\mu d\mu_{n,H} \\
&\geq \inf_{\theta \in \Theta_R} \inf_{\theta' \in Orb(\theta)} \int_{\Omega} \phi f_{\theta'} d\mu \\
&= \inf_{\theta \in \Theta_R} \int_{\Omega} \phi f_\theta d\mu.
\end{aligned}$$

Now we prove the second statement of the Corollary. First, note that

$$\begin{aligned}
\beta^*(v_\tau(\theta)) &= \sup_{\phi \in \Gamma(g,\gamma)} \int_{\Omega} \phi f_{v_\tau(\theta)} d\mu \\
&= \sup_{\left(\widetilde{\phi} \circ u_{\tau^{-1}}\right) \in \Gamma(g,\gamma)} \int_{\Omega} \widetilde{\phi} f_\theta d\mu \\
&= \beta^*(\theta),
\end{aligned}$$

because $(\phi \circ u_\tau) \in \Gamma(g,\gamma)$ if and only if $\phi \in \Gamma(g,\gamma)$, by previous results. Hence,

$$\begin{aligned}
\sup_{\theta \in \Theta_R} \left(\beta^*(\theta) - \int_{\Omega} \phi_I f_\theta d\mu\right) &= \sup_{\theta \in \Theta_R} \lim_{n \to \infty} \left(\beta^*(v_\tau(\theta)) - \int_G \int_{\Omega} \phi f_{v_\tau(\theta)} d\mu d\mu_{n,H}\right) \\
&\leq \sup_{\theta \in \Theta_R} \sup_{\theta' \in Orb(\theta)} \left(\beta^*(\theta') - \int_{\Omega} \phi f_{\theta'} d\mu\right) \\
&= \sup_{\theta \in \Theta_R} \left(\beta^*(\theta) - \int_{\Omega} \phi f_\theta d\mu\right),
\end{aligned}$$

and we conclude the Corollary. ■

4.3 Proofs of Chapter 3

Proof. [Proposition (17)] a) Since $\Gamma(g,\gamma) \subseteq \Gamma^n(g,\gamma)$ and $\{\theta^n_{R,i}\}_{i=1}^n \subseteq \Theta_R$ we have

$$\begin{aligned}
\sup_{\phi\in\Gamma(g,\gamma)} \inf_{\theta\in\Theta_R} \int_\Omega \phi f_\theta d\mu &\leq \sup_{\phi\in\Gamma^n(g,\gamma)} \inf_{\theta\in\Theta_R} \int_\Omega \phi f_\theta d\mu \\
&\leq \sup_{\phi\in\Gamma^n(g,\gamma)} \inf_{\theta\in\{\theta^n_{R,i}\}_{i=1}^n} \int_\Omega \phi f_\theta d\mu \\
&= \inf_{\theta\in\{\theta^n_{R,i}\}_{i=1}^n} \int_\Omega \phi^{n,*} f_\theta d\mu.
\end{aligned}$$

But from ii) and i)

$$\begin{aligned}
\inf_{\theta\in\{\theta^n_{R,i}\}_{i=1}^n} \int_\Omega \phi^{n,*} f_\theta d\mu &= \inf_{\theta\in\Theta_R} \int_\Omega \phi^{n,*} f_\theta d\mu \\
&\leq \sup_{\phi\in\Gamma(g,\gamma)} \inf_{\theta\in\Theta_R} \int_\Omega \phi f_\theta d\mu,
\end{aligned}$$

and we can conclude that $\phi^{n,*}$ is a GM test.

b) For the same reasons as a)

$$\inf_{\phi\in\Gamma^n(g,\gamma)} \sup_{\theta\in\{\theta^n_{R,i}\}_{i=1}^n} \left(\beta^*(\theta) - \int_\Omega \phi f_\theta d\mu\right) \leq \inf_{\phi\in\Gamma(g,\gamma)} \sup_{\theta\in\Theta_R} \left(\beta^*(\theta) - \int_\Omega \phi f_\theta d\mu\right),$$

but from ii) and i)

$$\sup_{\theta\in\Theta_R} \left(\beta^*(\theta) - \int_\Omega \phi^{n,\dagger} f_\theta d\mu\right) = \inf_{\phi\in\Gamma^n(g,\gamma)} \sup_{\theta\in\{\theta^n_{R,i}\}_{i=1}^n} \left(\beta^*(\theta) - \int_\Omega \phi f_\theta d\mu\right),$$

we conclude that $\phi^{n,\dagger}$ is a GMR test. ■

Proof. [Theorem (18)] a) Let $\phi^* \in \Gamma(g,\gamma)$ and $\left(\lambda^*, \left(\eta^*_{k,1}, \eta^*_{k,2}\right)_{k=1}^l\right) \in (\Lambda, \Xi)$ be solutions of the GMP and its Dual Problem. For each $n \in \mathbb{N}$ define

$$\lambda^{n,*}(\cdot) := \sum_{i=1}^n \lambda^*(\Pi^n_{R,i})\chi_{\{\theta^n_{R,i}\}}(\cdot), \tag{4.85}$$

$$\eta^{n,*}_{k,1}(\cdot) := \sum_{i=1}^n \eta^*_{k,1}(\Pi^n_{k,S,i})\chi_{\{\theta^n_{k,S,i}\}}(\cdot), \tag{4.86}$$

$$\eta^{n,*}_{k,2}(\cdot) := \sum_{i=1}^n \eta^*_{k,2}(\Pi^n_{k,S,i})\chi_{\{\theta^n_{k,S,i}\}}(\cdot), \tag{4.87}$$

$k = 1, ..., l$, sequence of measures defined on the Borel σ-algebra of Θ_R, $\Theta_{k,S}$ and $\Theta_{k,S}$, $k = 1, ..., l$, respectively. In the first part of the proof we prove that the sequences $(\lambda^{n,*})_{n=1}^\infty$, $\left(\eta^{n,*}_{k,1}\right)_{n=1}^\infty$ and $\left(\eta^{n,*}_{k,2}\right)_{n=1}^\infty$ weakly-star converges to $\lambda^*, \eta^*_{k,1}$ and $\eta^*_{k,2}$, $k = 1, ..., l$, respectively, as $||P^n_R||, ||P^n_S|| \to 0$. For any $h \in C(\Theta_R)$

$$|\int_{\Theta_R} h(\theta)d\lambda^{n,*} - \int_{\Theta_R} h(\theta)d\lambda^*| = |\int_{\Theta_R} h^n(\theta)d\lambda^* - \int_{\Theta_R} h(\theta)d\lambda^*|, \tag{4.88}$$

where $h^n(\theta) := \sum_{i=1}^n h(\theta^n_{R,i})1_{\{\Pi^n_{R,i}\}}(\theta)$. Since, for every $\theta \in \Theta_R$, $h^n(\theta) := \sum_{i=1}^n h(\theta^n_{R,i})1_{\{\Pi^n_{R,i}\}}(\theta) \to h(\theta)$ when $||P^n_R||| \to 0$ and $|h^n(\theta)| \leq ||h||_\infty$, we can apply the Dominated Convergence Theorem in (4.88) to obtain:

$$|\int_{\Theta_R} h(\theta)d\lambda^{n,*} - \int_{\Theta_R} h(\theta)d\lambda^*| \to 0,$$

when $||P_R^n||| \to 0$.Using the same arguments we prove that the sequences $\left(\eta_{k,1}^{n,*}\right)_{n=1}^{\infty}$ and $\left(\eta_{k,2}^{n,*}\right)_{n=1}^{\infty}$ weakly-star converges to $\eta_{k,1}^*$ and $\eta_{k,2}^*$, $k = 1, ..., l$, respectively. For the second part of the proof consider the set of measures

$$(\Lambda^n, \Xi^n) := \{\left(\lambda, \left(\eta_{k,1}, \eta_{k,2}\right)_{k=1}^l\right) \in (\Lambda, \Xi) : \lambda(\cdot) = \sum_{i=1}^n \underline{\lambda}_i \chi_{\{\theta_{R,i}^n\}}(\cdot);\ \eta_{k,1}(\cdot) = \sum_{i=1}^n \underline{\eta}_{k,1,i} \chi_{\{\theta_{k,S,i}^n\}}(\cdot);$$

$$\eta_{k,2}(\cdot) = \sum_{i=1}^n \underline{\eta}_{k,2,i} \chi_{\{\theta_{k,S,i}^n\}}(\cdot) \text{ where } \left(\underline{\lambda}_i, \left(\underline{\eta}_{k,1,i}, \underline{\eta}_{k,2,i}\right)_{k=1}^l\right)_{i=1}^n \in \mathbb{R}_+^{n(1+2l)}\}.$$

Since $(\Lambda^n, \Xi^n) \subseteq (\Lambda, \Xi)$ and $(\lambda^{n,*}, \left(\eta_{k,1}^{n,*}, \eta_{k,2}^{n,*}\right)_{k=1}^l) \in (\Lambda^n, \Xi^n)$ we obtain

$$\begin{aligned}
&\inf_{(\lambda,\left(\eta_{k,1},\eta_{k,2}\right)_{k=1}^l)\in(\Lambda,\Xi)} \sum_{k=1}^l \left(\int_{\Theta_{k,S}} \gamma_{k,2}(\theta) d\eta_{k,2} - \int_{\Theta_{k,S}} \gamma_{k,1}(\theta) d\eta_{k,1}\right) \\
&\qquad + \int_\Omega \left(\int_{\Theta_R} f_\theta d\lambda - \sum_{k=1}^l \int_{\Theta_{k,S}} g_\theta^k d(\eta_{k,2} - \eta_{k,1})\right)^+ d\mu \\
&\leq \inf_{(\lambda,\left(\eta_{k,1},\eta_{k,2}\right)_{k=1}^l)\in(\Lambda^n,\Xi^n)} \sum_{k=1}^l \left(\int_{\Theta_{k,S}} \gamma_{k,2}(\theta) d\eta_{k,2} - \int_{\Theta_{k,S}} \gamma_{k,1}(\theta) d\eta_{k,1}\right) \\
&\qquad + \int_\Omega \left(\int_{\Theta_R} f_\theta d\lambda - \sum_{k=1}^l \int_{\Theta_{k,S}} g_\theta^k d(\eta_{k,2} - \eta_{k,1})\right)^+ d\mu \\
&\leq \sum_{k=1}^l \left(\int_{\Theta_{k,S}} \gamma_{k,2}(\theta) d\eta_{k,2}^{n,*} - \int_{\Theta_{k,S}} \gamma_{k,1}(\theta) d\eta_{k,1}^{n,*}\right) \\
&\qquad + \int_\Omega \left(\int_{\Theta_R} f_\theta d\lambda^{n,*} - \sum_{k=1}^l \int_{\Theta_{k,S}} g_\theta^k d(\eta_{k,2}^{n,*} - \eta_{k,1}^{n,*})\right)^+ d\mu.
\end{aligned} \tag{4.89}$$

The last equation, however, as $||P_R^n||, ||P_S^n|| \to 0$ converges to

$$\begin{aligned}
&\sum_{k=1}^l \left(\int_{\Theta_{k,S}} \gamma_{k,2}(\theta) d\eta_{k,2}^{n,*} - \int_{\Theta_{k,S}} \gamma_{k,1}(\theta) d\eta_{k,1}^{n,*}\right) + \int_\Omega \left(\int_{\Theta_R} f_\theta d\lambda^{n,*} - \sum_{k=1}^l \int_{\Theta_{k,S}} g_\theta^k d(\eta_2^{n,*} - \eta_1^{n,*})\right)^+ d\mu \\
&\to \sum_{k=1}^l \left(\int_{\Theta_S} \gamma_2(\theta) d\eta_2^* - \int_{\Theta_S} \gamma_1(\theta) d\eta_1^*\right) + \int_\Omega \left(\int_{\Theta_R} f_\theta d\lambda^* - \sum_{k=1}^l \int_{\Theta_{k,S}} g_\theta^k d(\eta_{k,2}^* - \eta_{k,1}^*)\right)^+ d\mu,
\end{aligned} \tag{4.90}$$

the solution of the Dual Problem. Indeed, the first term convergence:

$$\sum_{k=1}^l \left(\int_{\Theta_{k,S}} \gamma_{k,2}(\theta) d\eta_{k,2}^{n,*} - \int_{\Theta_{k,S}} \gamma_{k,1}(\theta) d\eta_{k,1}^{n,*}\right) \to \sum_{k=1}^l \left(\int_{\Theta_S} \gamma_2(\theta) d\eta_2^* - \int_{\Theta_S} \gamma_1(\theta) d\eta_1^*\right),$$

holds because the measures $\left(\lambda^{n,*}, \left(\eta_{k,1}^{n,*}, \eta_{k,2}^{n,*}\right)_{k=1}^l\right)$ weakly-star converges to $\left(\lambda^*, \left(\eta_{k,1}^*, \eta_{k,2}^*\right)_{k=1}^l\right)$. For the convergence:

$$\int_\Omega \left(\int_{\Theta_R} f_\theta d\lambda^{n,*} - \sum_{k=1}^l \int_{\Theta_{k,S}} g_\theta^k d(\eta_{k,2}^{n,*} - \eta_{k,1}^{n,*})\right)^+ d\mu \to \int_\Omega \left(\int_{\Theta_R} f_\theta d\lambda^* - \sum_{k=1}^l \int_{\Theta_{k,S}} g_\theta^k d(\eta_{k,2}^* - \eta_{k,1}^*)\right)^+ d\mu,$$

note that for every $w \in \Omega$

$$\left(\int_{\Theta_R} f_\theta(w) d\lambda^{n,*} - \sum_{k=1}^l \int_{\Theta_{k,S}} g_\theta^k d(\eta_{k,2}^{n,*} - \eta_{k,1}^{n,*})\right)^+ \to \left(\int_{\Theta_R} f_\theta(w) d\lambda^* - \sum_{k=1}^l \int_{\Theta_{k,S}} g_\theta^k(w) d(\eta_{k,2}^* - \eta_{k,1}^*)\right)^+$$

and

$$\left(\int_{\Theta_R} f_\theta(w)d\lambda^{*,n} - \int_{\Theta_{k,S}} g^k_\theta(w)d(\eta^{*,n}_{k,2} - \eta^{*,n}_{k,1})\right)^+$$

$$\leq \sup_{\theta\in\Theta_R} |f_\theta(w)|\lambda^*(\Theta_R) + \sum_{k=1}^{l} \sup_{\theta\in\Theta_{k,S}} |g^k_\theta(w)| \left(\eta^*_{k,2}(\Theta_{k,S}) + \eta^*_{k,1}(\Theta_{k,S})\right), \tag{4.91}$$

where (4.91) is a μ-integrable function by *Assumption 5*. Hence we can apply the Dominated Convergence Theorem to obtain:

$$\int_\Omega \left(\int_{\Theta_R} f_\theta d\lambda^{n,*} - \sum_{k=1}^{l}\int_{\Theta_{k,S}} g^k_\theta d(\eta^{n,*}_{k,2} - \eta^{n,*}_{k,1})\right)^+ d\mu \to \int_\Omega \left(\int_{\Theta_R} f_\theta d\lambda^* - \int_{\Theta_{k,S}} g^k_\theta d(\eta^*_{k,2} - \eta^*_{k,1})\right)^+ d\mu.$$

Finally, we use the Strong Duality property of the Theorem (12) in the n-GMMP and GMMP to obtain

$$\begin{aligned}
&\inf_{\theta\in\{\theta^n_{R,i}\}_{i=1}^n} \int_\Omega \phi^{n,*} f_\theta d\mu \\
&= \sup_{\phi\in\Gamma^n(g,\gamma)} \inf_{\theta\in\{\theta^n_{R,i}\}_{i=1}^n} \int_\Omega \phi f_\theta d\mu \\
&= \inf_{\left(\underline{\lambda}, \left(\underline{\eta}_{k,1}, \underline{\eta}_{k,2}\right)_{k=1}^l\right)\in\mathcal{S}} \sum_{i=1}^{n}\sum_{k=1}^{l} \left(\gamma_{k,2}(\theta^n_{k,S,i})\underline{\eta}_{k,2,i} - \gamma_{k,1}(\theta^n_{k,S,i})\underline{\eta}_{k,1,i}\right) \\
&\qquad\qquad + \int_\Omega \left(\sum_{i=1}^{n}\left(f_{\theta^n_{R,i}}\underline{\lambda}_i - \sum_{k=1}^{l} g^k_{\theta^n_{k,S,i}}(\underline{\eta}_{k,2,i} - \underline{\eta}_{k,1,i})\right)\right)^+ d\mu \\
&\to \inf_{\left(\lambda, \left(\eta_{k,1}, \eta_{k,2}\right)_{k=1}^l\right)\in(\Lambda,\Xi)} \sum_{k=1}^{l} \left(\int_{\Theta_{k,S}} \gamma_{k,2}(\theta)d\eta_{k,2} - \int_{\Theta_{k,S}} \gamma_{k,1}(\theta)d\eta_{k,1}\right) \\
&\qquad\qquad + \int_\Omega \left(\int_{\Theta_R} f_\theta d\lambda - \sum_{k=1}^{l}\int_{\Theta_{k,S}} g^k_\theta d(\eta_{k,2} - \eta_{k,1})\right)^+ d\mu \\
&= \sup_{\phi\in\Gamma(g,\gamma)} \inf_{\theta\in\Theta_R} \int_\Omega \phi f_\theta d\mu \\
&= \inf_{\theta\in\Theta_R} \int_\Omega \phi^* f_\theta d\mu,
\end{aligned}$$

where $\mathcal{S} = \{\left(\underline{\lambda}, \left(\underline{\eta}_{k,1}, \underline{\eta}_{k,2}\right)_{k=1}^l\right) \in \mathbb{R}_+^{n(1+2l)} : \sum_{i=1}^{n}\underline{\lambda}_i = 1\}$. For part b), let $\phi^\dagger \in \Gamma(g,\gamma)$ and $(\lambda^\dagger, \left(\eta^\dagger_{k,1}, \eta^\dagger_{k,2}\right)_{k=1}^l) \in (\Lambda,\Xi)$ be solutions of the GMR Problem (14) and its Dual Problem (2.22). Define the sequence $(\lambda^{n,\dagger}, \left(\eta^{n,\dagger}_{k,1}, \eta^{n,\dagger}_{k,2}\right)_{k=1}^l)_{n=1}^\infty$ as in (4.85),(4.86),(4.87) using $(\lambda^\dagger, \left(\eta^\dagger_{k,1}, \eta^\dagger_{k,2}\right)_{k=1}^l)$. For the same reasons of Theorem (18) proof's, the sequence of measures $(\lambda^{n,\dagger}, \left(\eta^{n,\dagger}_{k,1}, \eta^{n,\dagger}_{k,2}\right)_{k=1}^l)$ weakly-star converges to $(\lambda^\dagger, \left(\eta^\dagger_{k,1}, \eta^\dagger_{k,2}\right)_{k=1}^l)$ as $||P^n_R||, ||P^n_S|| \to 0$.

Since $(\Lambda^n,\Xi^n)\subseteq(\Lambda,\Xi)$ and $(\lambda^{n,\dagger},\left(\eta_{k,1}^{n,\dagger},\eta_{k,2}^{n,\dagger}\right)_{k=1}^{l})\in(\Lambda^n,\Xi^n)$,

$$
\begin{aligned}
& \sup_{(\lambda,\left(\eta_{k,1},\eta_{k,2}\right)_{k=1}^{l})\in(\Lambda,\Xi)} \int_{\Theta_R}\beta^*(\theta)\ d\lambda+\sum_{k=1}^{l}\left(\int_{\Theta_{k,S}}\gamma_{k,1}(\theta)d\eta_{k,1}-\int_{\Theta_{k,S}}\gamma_{k,2}(\theta)d\eta_{k,2}\right)\\
&\qquad -\int_{\Omega}\left(\int_{\Theta_R}f_\theta d\lambda-\sum_{k=1}^{l}\int_{\Theta_{k,S}}g_\theta^k d(\eta_{k,2}-\eta_{k,1})\right)^+ d\mu\\
\geq\ & \sup_{(\lambda,\left(\eta_{k,1},\eta_{k,2}\right)_{k=1}^{l})\in(\Lambda^n,\Xi^n)} \int_{\Theta_R}\beta^*(\theta)d\lambda+\sum_{k=1}^{l}\left(\int_{\Theta_{k,S}}\gamma_{k,1}(\theta)d\eta_{k,1}-\int_{\Theta_{k,S}}\gamma_{k,2}(\theta)d\eta_{k,2}\right)\\
&\qquad -\int_{\Omega}\left(\int_{\Theta_R}f_\theta d\lambda-\sum_{k=1}^{l}\int_{\Theta_{k,S}}g_\theta^k d(\eta_{k,2}-\eta_{k,1})\right)^+ d\mu\\
\geq\ & \int_{\Theta_R}\beta^*(\theta)d\lambda^{n,\dagger}+\sum_{k=1}^{l}\left(\int_{\Theta_{k,S}}\gamma_{k,1}(\theta)d\eta_{k,1}^{n,\dagger}-\int_{\Theta_{k,S}}\gamma_{k,2}(\theta)d\eta_{k,2}^{n,\dagger}\right) \qquad (4.92)\\
&\qquad -\int_{\Omega}\left(\int_{\Theta_R}f_\theta d\lambda^{n,\dagger}-\sum_{k=1}^{l}\int_{\Theta_{k,S}}g_\theta^k d(\eta_{k,2}^{n,\dagger}-\eta_{k,1}^{n,\dagger})\right)^+ d\mu \qquad (4.93)\\
\rightarrow\ & \int_{\Theta_R}\beta^*(\theta)d\lambda^{*}+\sum_{k=1}^{l}\left(\int_{\Theta_{k,S}}\gamma_{k,1}(\theta)d\eta_{k,1}^{*}-\int_{\Theta_{k,S}}\gamma_{k,2}(\theta)d\eta_{k,2}^{*}\right)\\
&\qquad -\int_{\Omega}\left(\int_{\Theta_R}f_\theta d\lambda^{*}-\sum_{k=1}^{l}\int_{\Theta_{k,S}}g_\theta^k d(\eta_{k,2}^{*}-\eta_{k,1}^{*})\right)^+ d\mu \qquad (4.94)
\end{aligned}
$$

when $||P_R^n||,||P_S^n||\to 0$, where the convergence in (4.94) follows by the weakly-star convergence of $\left(\lambda^{n,*},\eta_1^{n,*},\eta_2^{n,*}\right)$ to $(\lambda^*,\eta_1^*,\eta_2^*)$ and the Dominated Convergence Theorem. Hence,

$$
\begin{aligned}
& \sup_{\theta\in\{\theta_{R,i}^n\}_{i=1}^n}\left(\beta^*(\theta)-\int_{\Omega}\phi^{n,\dagger}f_\theta d\mu\right)\\
=\ & \inf_{\phi\in\Gamma^n(g,\gamma,\delta)}\ \sup_{\theta\in\{\theta_{R,i}^n\}_{i=1}^n}\left(\beta^*(\theta)-\int_{\Omega}\phi f_\theta d\mu\right)\\
=\ & \sup_{(\lambda,\left(\eta_{k,1},\eta_{k,2}\right)_{k=1}^{l})\in(\Lambda^n,\Xi^n)} \int_{\Theta_R}\beta^*(\theta)d\lambda+\sum_{k=1}^{l}\left(\int_{\Theta_{k,S}}\gamma_{k,1}(\theta)d\eta_{k,1}-\int_{\Theta_{k,S}}\gamma_{k,2}(\theta)d\eta_{k,2}\right)\\
&\qquad -\int_{\Omega}\left(\int_{\Theta_R}f_\theta d\lambda-\sum_{k=1}^{l}\int_{\Theta_{k,S}}g_\theta^k d(\eta_{k,2}-\eta_{k,1})\right)^+ d\mu
\end{aligned}
$$

$$
\begin{aligned}
&= \sup_{(\underline{\lambda},\underline{\eta}_1,\underline{\eta}_2)\in\mathcal{S}} \sum_{i=1}^{n} \left(\beta^*(\theta^n_{R,i})\underline{\lambda}_i + \sum_{k=1}^{l} \Big(\gamma_{k,1}(\theta^n_{k,S,i})\underline{\eta}_{k,1,i} - \gamma_{k,2}(\theta^n_{k,S,i})\underline{\eta}_{k,2,i} \Big) \right) \\
&\qquad\qquad - \int_\Omega \left(\sum_{i=1}^{n} \left(f_{\theta^n_{R,i}}\underline{\lambda}_i - \sum_{k=1}^{l} g^k_{\theta^n_{k,S,i}}(\underline{\eta}_{k,2,i} - \underline{\eta}_{k,1,i}) \right) \right)^+ d\mu \\
&\to \sup_{(\lambda,\eta_1,\eta_2)\in(\Lambda,\Xi)} \int_{\Theta_R} \beta^*(\theta) d\lambda + \sum_{k=1}^{l} \left(\int_{\Theta_{k,S}} \gamma_{k,1}(\theta) d\eta_{k,1} - \int_{\Theta_{k,S}} \gamma_{k,2}(\theta) d\eta_{k,2} \right) \\
&\qquad\qquad - \int_\Omega \left(\int_{\Theta_R} f_\theta d\lambda - \sum_{k=1}^{l} \int_{\Theta_{k,S}} g^k_\theta d(\eta_{k,2} - \eta_{k,1}) \right)^+ d\mu \\
&= \inf_{\phi\in\Gamma(g,\gamma)} \sup_{\theta\in\Theta_R} \left(\beta^*(\theta) - \int_\Omega \phi f_\theta d\mu \right) \\
&= \sup_{\theta\in\Theta_R} \left(\beta^*(\theta) - \int_\Omega \phi^\dagger f_\theta d\mu \right),
\end{aligned}
$$

when $||P^n_R||, ||P^n_S|| \to 0$. ■

Bibliography

Anderson, E., and P. Nash (1987): *Linear Programming in Infinite-Dimensional Spaces*. Willey.

Anderson, T. W., and H. Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46–63.

Andrews, D. W. K. (2012): "Similar-On-The-Boundary Tests For Moment Inequality Exists, But Have Poor Power," *Discussion Paper Yale University*, 1815.

Andrews, D. W. K., and P. J. Barwick (2012): "Inference For Parameters Defined By Moments Inequalities: A Recommended Moment Selection Procedure," *Econometrica*, 80, 2805–2826.

Andrews, D. W. K., X. Cheng, and P. Guggenberger (2011): "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests," *Cowles Foundation*, 1813.

Andrews, D. W. K., and P. Guggenberger (2010): "Applications of subsampling, hybrid, and size-correction methods," *Journal of Econometrics*, 158, 285–305.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2004): "Optimal Invariant Similar Tests for Instrumental Variables Regression," *NBER Working Paper*, p. t0299.

——— (2006a): "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74, 715–752.

——— (2006b): "Supplement to 'Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression'," *Econometrica Supplementary Material*, 74.

——— (2007): "Performance of Conditional Wald Tests in IV Regression with Weak Instruments," *Journal of Econometrics*, 139, 116–132.

——— (2008): "Optimal Two-Sided Nonsimilar Invariant Tests in IV Regression with Weak Instruments," *Journal of Econometrics*, 146, 241–254.

Andrews, D. W. K., and G. Soares (2007): "Inference For Parameters Defined By Moment Inequalities Using Generalized Moment Selection," *Cowles Foundation*, p. 1631.

Andrews, D. W. K., and J. H. Stock (2007): *'Inference with Weak Instruments' in Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, vol. 3. Cambridge University Press, Cambridge.

Angrist, J., and A. B. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?," *The Quarterly Journal of Economics*, 106, 979–1014.

Bauer, H. (2001): *Measure and Integration Theory*. W. de Gruyter, Berlin.

Bourbaki, N. (2004): *Elements of Mathematics Integration II*. Springer-Verlag, Berlin Heidelberg.

Brezis, H. (2011): *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext.

Cattaneo, M., R. Crump, and M. Jansson (2012): "Optimal Inference for Instrumental Variables Regression with Non-Gaussian Errors," *Journal of Econometrics*, 167, 1–15.

Chamberlain, G. (2007): "Decision Theory Applied To an Instrumental Variables Model," *Econometrica*, 75, 609–652.

Chernozhukov, V., C. Hansen, and M. Jansson (2009): "Admissible Invariant Similar Tests for Instrumental Variables Regression," *Econometric Theory*, 25, 806–818.

Chiburis, R. C. (2009): "Approximately Most Powerful Tests For Moment Inequalities," *Working Paper Princeton University*.

Conway, J. B. (1990): *A Course in Functional Analysis*. Springer-Verlang, New York.

Cruz, L. M., and M. J. Moreira (2005): "On The Validity Of Econometric Techiniques With Weak Instruments: Inference On Returns To Education Using Compulsory School Attendance Laws," *Journal of Human Resources*, 40, 393–410.

Cvitanic, J., and I. Karatzas (2001): "Generalized Neyman Pearson Lemma Via Convex Duality," *Bernoulli*, 7, 79–97.

Dantzig, G. B. (1963): *Linear Programming And Extensions*. Princeton University Press, New Jersey.

Donald, S. G., and W. K. Newey (2001): "Choosing the Number of Instruments," *Econometrica*, 69, 1161–1192.

Dufour, J. M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1388.

——— (2003): "Presidential Address: Identification, Weak Instruments, and Statistical Inference in Econometrics," *Canadian Journal of Economics*, 36, 767–808.

Ekeland, I., and R. Teman (1999): *Convex analysis and variational problems*. Springer Society for Industrial and Applied Mathematics.

Fuller, W. A. (1977): "Some properties of a modification of the limited information estimator," *Econometrica*, 45, 939–954.

Gushchin, A. (2015): "A Characterization Of Maximin Tests For Two Composite Hypotheses," *Mathematical Methods of Statistics*, 24, 110–121.

Imbens, G. W., and C. F. Manski (2004): "Confidence Intervals Of Partially Identified Parameters," *Econometrica*, 72, 1845–1857.

Kerstan, J., and K. Matthes (1969): "Gleichverteilungseigenschaften von Faltungen von Verteilungsgesetzen auf lokal-kompakten Abelschen Gruppen II," *Math. Nachr.*, 41, 121–132.

Kleibergen, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70, 1781–1803.

Krafft, O., and H. Witting (1967): "Optimale Tests und ungiinstigste Verteilungen. Z.," *Wahrscheinlichkeitstheorie Verw. Geb.*, 7, 289–302.

LeCam, L. (1986): *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlang, New York.

Lehamnn, E. L. (1952): "On The Existence Of Least-Favorable Distributions," *The Annals of Mathematical Statistics*, 23, 408–416.

Lehmann, E. L., and J. Romano (2005): *Testing Statistical Hypotheses*. Springer Verlang, New York.

Mills, B., M. J. Moreira, and L. P. Vilela (2014): "Tests Based On t-Statistics For IV Regression With Weak Instruments," *Journal of Econometrics*, 182, 351–363.

Moreira, H., and M. J. Moreira (2013): "Contribution To The Theory Of Optimal Tests," *Ensaios Economicos*, 747.

Moreira, M. J. (2002): "Tests with Correct Size in the Simultaneous Equations Model," Ph.D thesis, Berkeley.

——— (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71, 1027–1048.

——— (2009): "A Maximum Likelihood Method for the Incidental Parameter Problem," *Annals of Statistics*, 37, 3660–3696.

Nagar, A. (1959): "The Bias and Moment Matrix of the General k-Class Estimators of Parameters in Simultaneous Equations," *Econometrica*, 65, 557–586.

Reinhardt, H. (1961): "The Use Of Least-Favorable Distributions in Testing Statistical Hypotheses," *The Annals of Mathematical Statistics*, 32, 1034–1041.

Rockafellar, R. T. (1974): *Conjugate Duality and Optimization*. Regional Conference Series in Applied Mathematics.

Rosen, A. M. (2007): "Confidence Sets For Partially Identified Parameters That Satisfy A Finite Number Of Moment Inequalities," *Working Paper University College London.*

Rudloff, B., and I. Karatzas (2010): "Testing Composite Hypotheses Via Convex Duality," *Bernoulli*, 16, 1224–1239.

Savage, L. (1951): "The Theory of Statistical Decision," *Journal of the American Statistical Association*, 46, 55–67.

Schaafsma, W. (1951): "Minimax Risk and Unbiasedness for Multiple Most Problems of Type I," *The Annals of Mathematical Statistics*, 40, 1684–1720.

——— (1966): "Hypothesis Testing Problems With The Alternative Restricted By A Number Of Inequalities," Thesis, Noordhoff, Groningen.

——— (1968): "A Comparison of the Most Stringent and the Most Stringent Somewhere Most Powerful Test for Certain Problems with Restricted Alternative," *The Annals of Mathematical Statistics*, 39, 531–546.

Schaafsma, W., and L. J. Smid (1966): "Most Stringent Somewhere Most Powerfull Tests Against Alternative Hypothesis Restricted by a Number of Linear Inequalities," *The Annals of Mathematical Statistics*, 37, 1161–1172.

Sion, M. (1958): "On General Minimax Theorems," *Pacific Journal of Mathematics*, 8, 171–176.

Staiger, D., and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 556–586.

Stock, J. H., J. Wright, and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20, 518–529.

Strasser, H. (1985): *Mathematical Theory of Statistics*. W. de Gruyter.

Tsai, M., and P. K. Sen (1993): "On The Local Optimality Of Optimal Linear Tests For Restricted Alternatives," *Statistica Sinica*, 3, 103–115.

Vajda, I. (1989): *Theory of Statistical Inference and Information*. Dordrecht, Netherlands, Kluwer.

Wald, A. (1945): "Statistical Decision Functions Which Minimize The Maximum Risk,," *Annals of Mathematics*, 46, 265–280.

Witting, H. (1985): *Mathematische Statistik I*. Stuttgart, FRG, Teubner.

Yosida, K. (1995): *Functional Analysis*. Classics in Mathematics.