ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

Econometrica, Vol. 84, No. 6 (November, 2016), 2155–2182

CONDITIONAL LINEAR COMBINATION TESTS FOR WEAKLY IDENTIFIED MODELS

ISAIAH ANDREWS
Massachusetts Institute of Technology, Cambridge, MA 02142, U.S.A.

Econometrica, Vol. 84, No. 6 (November, 2016), 2155–2182

CONDITIONAL LINEAR COMBINATION TESTS FOR WEAKLY IDENTIFIED MODELS

BY ISAIAH ANDREWS[1]

We introduce the class of conditional linear combination tests, which reject null hypotheses concerning model parameters when a data-dependent convex combination of two identification-robust statistics is large. These tests control size under weak identification and have a number of optimality properties in a conditional problem. We show that the conditional likelihood ratio test of Moreira (2003) is a conditional linear combination test in models with one endogenous regressor, and that the class of conditional linear combination tests is equivalent to a class of quasi-conditional likelihood ratio tests. We suggest using minimax regret conditional linear combination tests and propose a computationally tractable class of tests that plug in an estimator for a nuisance parameter. These plug-in tests perform well in simulation and have optimal power in many strongly identified models, thus allowing powerful identification-robust inference in a wide range of linear and nonlinear models without sacrificing efficiency if identification is strong.

KEYWORDS: Instrumental variables, nonlinear models, power, size, test, weak identification.

1. INTRODUCTION

RESEARCHERS IN ECONOMICS ARE FREQUENTLY INTERESTED IN inference on causal or structural parameters. Unfortunately, in cases where the data contain only limited information useful for estimating these parameters, commonly used approaches to estimation and inference can break down and researchers who rely on such techniques risk drawing highly misleading inferences. Models where the usual approaches to inference fail due to limited information about model parameters are referred to as *weakly identified*. A large and growing literature develops identification-robust hypothesis tests, which control size regardless of identification strength and so limit the probability of rejecting true hypotheses in weakly identified contexts. The results to date on the power of

[1]The author is grateful to Anna Mikusheva, Whitney Newey, and Jerry Hausman for their guidance and support, and to Don Andrews, Arun Chandrasekhar, Victor Chernozhukov, Denis Chetverikov, Kirill Evdokimov, Benjamin Feigenberg, Patricia Gomez-Gonzalez, Patrik Guggenberger, Bruce Hansen, Sally Hudson, Peter Hull, Conrad Miller, Scott Nelson, Jose Montiel Olea, Miikka Rokkanen, Adam Sacarny, Annalisa Scognamiglio, Brad Shapiro, Ashish Shenoy, Stefanie Stantcheva, the participants of the MIT Econometrics Lunch and the Chicago Econometrics working group, and seminar participants at Brown, Cambridge, Colegio Carlo Alberto, Columbia, Cornell, FGV, INSEAD, Northwestern, NYU, Penn State, Princeton, Wisconsin, Yale, and the 2014 Winter meetings of the Econometric Society for helpful comments. The author thanks Marcelo Moreira for help in finding and resolving a problem with the implementation of the Moreira and Moreira (2015) tests in an earlier draft, and for productive discussions about the Moreira and Moreira (2015) procedures. NSF Graduate Research Fellowship support under Grant 1122374 is gratefully acknowledged.

 DOI: 10.3982/ECTA12407

identification-robust tests are, however, quite limited. In this paper, we develop powerful identification-robust tests applicable to a wide range of models. Our approach relies on two innovations. First, we introduce a novel class of procedures, the class of conditional linear combination tests, which includes many known robust tests. Second, we suggest choosing conditional linear combination tests that minimize maximum regret, which is an intuitive optimality criterion not previously applied in this setting.

We consider tests based on the generalized Anderson–Rubin (S) statistic introduced by Stock and Wright (2000) and the score (K) and conditioning (D) statistics introduced by Kleibergen (2005). Tests based on S have stable power but are inefficient under strong identification, while tests based on K are efficient when identification is strong but can have low power when identification is weak. In many models, D can be viewed as measuring identification strength, and its behavior governs the performance of tests based on K. Tests based on these three statistics (or their analogs for generalized minimum distance, generalized empirical likelihood, or other settings) comprise the bulk of procedures which have been studied in the weak-identification literature to date.

We show that, conditional on D, tests which reject when convex combinations of the S and K statistics are large are admissible, locally most powerful against particular sequences of alternatives, and weighted average power maximizing for a continuum of different weight functions. Motivated by these facts, we propose the class of *conditional linear combination* (CLC) tests, which use information from D to determine how to weight the S and K statistics, and select critical values based on D in such a way that all tests in this class have correct size. Further, all CLC tests are unbiased, in the sense that their rejection probability under any alternative is at least as high as their rejection probability under the null. The class of conditional linear combination tests is large, and includes the S test of Stock and Wright (2000) and K test of Kleibergen (2005) for GMM and the conditional likelihood ratio (CLR) test of Moreira (2003) for linear instrumental variables (IV) models with a single endogenous regressor. More generally, we prove that the class of CLC tests is equivalent to a suitably defined class of quasi-CLR tests. All CLC tests are unbiased, so one implication of this result is that all quasi-CLR tests are unbiased as well. This is to our knowledge a new result even for the CLR test in linear IV with homoscedastic errors and a single endogenous regressor.

Our second innovation is to use minimax regret CLC tests. This approach selects CLC tests with power functions as close as possible to the power envelope for this class in a uniform sense. By construction, these tests minimize the largest margin by which the power of the test selected could fall short relative to any other CLC test the researcher might have picked, thus minimizing the extent to which a researcher might regret their choice. Minimax regret has recently seen use in other areas of economics and econometrics (see Stoye (2009) for references) but has not to our knowledge been applied to the problem of selecting powerful tests for weakly identified models. Minimax regret tests must be obtained numerically which, while quite straightforward for some models, can be computationally daunting for others. In contexts where calculating true minimax regret tests is infeasible, we suggest a class of computationally simple plug-in minimax regret tests that plug in an estimate for a nuisance parameter.

We show that our plug-in tests perform well in linear IV. Specifically, in linear IV with homoscedastic Gaussian errors and one endogenous regressor, we show that plug-in minimax regret tests using reasonable plug-in estimators match the near-optimal performance of the CLR test established by D. Andrews, Moreira, and Stock (2006, henceforth AMS). Given that much of the data encountered in econometric practice is dependent (serially or spatially correlated, clustered), heteroscedastic, or both, however, it is of considerable interest to examine the performance of weak instrument-robust tests more broadly. To this end, we calibrate a simulation to match heteroscedastic time-series data used by Yogo (2004) and find that our plug-in minimax regret test substantially outperforms Kleibergen's (2005) quasi-CLR test for general GMM models. We further find that our approach offers power competitive with the conditional QLR test of I. Andrews and Mikusheva (2016a) and the weighted average power optimal MM1-SU and MM2-SU tests of Moreira and Moreira (2015, henceforth MM). The under-performance of Kleibergen's quasi-CLR test can be traced to the fact that the K statistic may perform especially poorly in non-homoscedastic IV. Kleibergen's test uses the CLR weight function, which is optimal under homoscedasticity but does not account for deterioration in the performance of the K statistic when we move away from the homoscedastic case. In contrast, the plug-in test proposed in this paper successfully accounts for the covariance structure of the data and delivers powerful, stable performance in both the homoscedastic and non-homoscedastic cases.

We consider inference on parameters in linear IV and minimum distance models as recurring examples. Similarly to Müller (2011), we assume that certain functions of the data converge in distribution to random variables in a limit problem and use this limit problem to study the performance of different procedures. To formally justify this approach, in the Supplement we derive a number of asymptotic results, showing that the asymptotic size and power of CLC tests under the assumed convergence are simply their size and power in the limit problem. We further show that a large class of CLC tests control size uniformly in heteroscedastic linear IV with a single endogenous regressor. Moreover, we give conditions under which CLC tests, and plug-in minimax regret tests in particular, will be asymptotically efficient under strong identification, in the sense of being asymptotically uniformly most powerful in classes of tests depending on (S, K, D). Applying these results to our examples, we show that the tests we propose are asymptotically efficient in linear IV and minimum distance models when identification is strong.

Before proceeding, it is worth relating the approach taken in this paper to the recent econometric literature on optimal testing in non-standard models, including Müller (2011), Elliott, Müller, and Watson (2015), Montiel-Olea

(2016), and MM. The approaches studied in those papers apply under a weak convergence condition like the one we assume, and in each case the authors derived tests maximizing weighted average power. If a researcher has a well-defined weight function over the alternative with respect to which they want to maximize average power, these approaches deliver optimal tests, either over the class of all tests or over the class of tests satisfying some auxiliary restrictions, and have a great deal to recommend them. In general, these tests are not available in closed form and will depend on the weight function chosen, however, and the nature of this dependence in a given context can be quite opaque. Consequently, in cases where the researcher has no particular weight function in mind, it can be unclear what a given choice of weight function will imply for the power of the resulting test. Indeed, as MM showed in their linear IV simulations, weighted average power optimal tests may sometimes have low power over empirically relevant regions of the parameter space. MM addressed this issue by restricting attention to classes of locally unbiased tests (their LU and SU tests).[2] Here, we take a different approach and adopt a minimax regret perspective which attempts to pick tests that lie as close as possible to the power envelope for the class of CLC tests. Relative to the papers discussed above, the approach of this paper greatly restricts the class of tests considered, first in confining attention to tests that depend only on S, K, and D, and then in further focusing on CLC tests. While this restriction reduces the strength of optimality statements, it renders the resulting tests much more transparent: conditional on D, the procedures discussed in this paper are simply tests based on a known convex combination of the S and K statistics, making it simple to understand their behavior. For example, it is easy to show that CLC tests are unbiased, while determining whether a given weighted average power optimal test is unbiased is typically a challenging exercise. This transparency has other advantages, and it is relatively straightforward to give conditions under which CLC tests will be efficient under strong identification. This is particularly true of plug-in minimax regret tests which, while not generally optimal from a minimax regret perspective, yield easy-to-characterize behavior under strong-identification asymptotics. In contrast, weighted average power optimal tests need not be efficient under strong identification, though Elliott, Müller, and Watson (2015) and Moreira and Moreira (2015) suggested particular tests which they showed are efficient under strong identification.

In the next section, we outline the weak convergence assumption that will form the basis of our analysis and illustrate this assumption using our IV and minimum distance examples. In Section 3, we define several statistics including S, K, and D and discuss tests which have been proposed based on these statistics. Section 4 considers the testing problem conditional on D, characterizes the class of tests based on (S, K, D) which are admissible in this conditional

[2]An earlier paper, Moreira and Moreira (2010), discussed the issue of approximating weighted average power optimal similar tests of a given size, but did not discuss the IV example.

problem, and shows that tests based on linear combinations of S and K are conditionally locally most powerful and weighted average power maximizing. Section 5 defines CLC tests and proves the equivalence of the class of CLC tests and a class of quasi-CLR test. Section 6 defines minimax regret CLC tests and plug-in tests. Section 7 shows that suitably defined plug-in minimax regret tests match the near-optimal performance of the CLR test under homoscedasticity and are competitive with existing alternatives in simulations calibrated to Yogo's (2004) data. Asymptotic results and all proofs may be found in the Supplement (I. Andrews (2016)). An empirical application, discussion of implementation, details of our examples, and additional simulation results may be found in the Supplemental Material, which is included with the replication files. To illustrate the application of our approach to a nonlinear example, we also apply our results to a generalized minimum distance approach to inference on new Keynesian Phillips curve parameters studied in Magnusson and Mavroeidis (2010).

2. WEAKLY IDENTIFIED LIMIT PROBLEMS

In this section, we describe a class of limit problems that arise in many weakly identified contexts and illustrate this class with two examples. We assume a sequence of models indexed by sample size T, where sample T has distribution $F_T(\theta, \gamma)$ for $\theta \in \Theta$ a p-dimensional parameter of interest and $\gamma \in \Gamma$ an l-dimensional consistently estimable nuisance parameter. We will be concerned with testing $H_0 : \theta = \theta_0$ and assume we observe three objects: a $k \times 1$ vector $g_T(\theta_0)$ which will typically be an appropriately scaled moment vector or distance function, a $k \times p$ matrix $\Delta g_T(\theta_0)$ which will often be some transformation of the Jacobian of $g_T(\theta)$ with respect to θ, and an estimate $\hat{\gamma}$ for γ. We assume that for all fixed $(\theta, \gamma) \in \Theta \times \Gamma$ we have

$$\begin{pmatrix} g_T(\theta_0) \\ \Delta g_T(\theta_0) \end{pmatrix} \to_d \begin{pmatrix} g \\ \Delta g \end{pmatrix} \tag{1}$$

and $\hat{\gamma} \to_p \gamma$ under the sequence of data-generating processes $F_T(\theta, \gamma)$, where

$$\begin{pmatrix} g \\ \operatorname{vec}(\Delta g) \end{pmatrix} \sim N\left(\begin{pmatrix} m \\ \operatorname{vec}(\mu) \end{pmatrix}, \begin{pmatrix} I & \Sigma_{g\theta} \\ \Sigma_{\theta g} & \Sigma_{\theta\theta} \end{pmatrix} \right), \tag{2}$$

and $m = m(\theta, \theta_0, \gamma) \in \mathcal{M}(\mu, \gamma)$ for a set $\mathcal{M}(\mu, \gamma) \subseteq \mathbb{R}^k$ which may depend on $\mu \in \mathbb{M}$ and γ. Here we use $\operatorname{vec}(A)$ to denote vectorization, which maps the $k \times p$ matrix A to a $kp \times 1$ vector. We further assume that $\Sigma_{\theta g}$ and $\Sigma_{\theta\theta}$ are continuous functions of γ and are thus consistently estimable. We will generally suppress the dependence of the terms in the limit problem on the parameters (θ, γ), writing simply m, μ, and so forth. We are interested in problems where the null hypothesis $\theta = \theta_0$ implies $m = 0$, and will focus on testing $H_0 : m = 0$, $\mu \in \mathbb{M}$ against $H_1 : m \in \mathcal{M}(\mu) \setminus \{0\}$, $\mu \in \mathbb{M}$.

Limit problems of the form (2) arise in a wide variety of weakly identified models. In the remainder of this section, we show that weakly identified instrumental variables and minimum distance models generate limit problems of this form, deferring some derivations to the Supplemental Material. In the Supplemental Material, we also show that weakly identified GMM models give rise to limiting problems of the form (2).

EXAMPLE I—Weak IV: Consider a linear instrumental variables model with a single endogenous regressor, written in reduced form,

$$Y = Z\pi\beta + V_1, \qquad X = Z\pi + V_2, \tag{3}$$

for Z a $T \times k$ matrix of instruments, X a $T \times 1$ vector of endogenous regressors, Y a $T \times 1$ vector of outcome variables, and V_1 and V_2 both $T \times 1$ vectors of residuals. We are interested in testing a hypothesis $H_0 : \beta = \beta_0$ about the scalar coefficient β. As elsewhere in the literature (see, e.g., AMS), we can accommodate additional exogenous regressors, but omit such variables here to simplify the exposition.

The identifying assumption in IV models is that $E[V_{1,t}Z_t] = E[V_{2,t}Z_t] = 0$ for Z_t the transpose of row t of Z, which allows us to view linear IV as a special case of GMM with moment condition

$$f_t(\beta) = (Y_t - X_t\beta)Z_t \tag{4}$$

and identifying assumption $E_\beta[f_t(\beta)] = 0$ (where $E_\theta[X]$ denotes the expectation of X under true parameter value θ). For fixed $\pi \neq 0$, it is straightforward to construct consistent, asymptotically normal GMM estimates based on (4) and to use these estimates to test hypotheses about β. The standard asymptotic approximations to the distribution of estimators and test statistics may, however, be quite poor if π is small relative to the sample size. To derive better approximations for this weakly identified case, following Staiger and Stock (1997) we can model the first-stage parameter π as changing with the sample size, taking $\pi_T = \frac{c}{\sqrt{T}}$ for a fixed vector $c \in \mathbb{R}^k$.

To derive the limit problem (2) for this model, define $f_T(\beta) = \frac{1}{T}\sum f_t(\beta)$ and let Ω be the asymptotic variance matrix of $\sqrt{T}(f_T(\beta_0)', -\frac{\partial}{\partial\beta}f_T(\beta_0)')'$,

$$\Omega = \begin{pmatrix} \Omega_{ff} & \Omega_{f\beta} \\ \Omega_{\beta f} & \Omega_{\beta\beta} \end{pmatrix} = \lim_{T\to\infty} \operatorname{Var}\left(\sqrt{T}\begin{pmatrix} f_T(\beta_0) \\ -\dfrac{\partial}{\partial\beta}f_T(\beta_0) \end{pmatrix}\right). \tag{5}$$

We assume that Ω_{ff} is full-rank. For $\hat{\Omega}$ a consistent estimator of Ω, define $g_T(\beta) = \sqrt{T}\hat{\Omega}_{ff}^{-\frac{1}{2}} f_T(\beta)$, $\Delta g_T(\beta) = -\sqrt{T}\hat{\Omega}_{ff}^{-\frac{1}{2}} \frac{\partial}{\partial\beta} f_T(\beta)$, and $\hat{\gamma} = \operatorname{vec}(\hat{\Omega})$. For $\theta =$

β, $\Theta = \mathbb{R}$, $\gamma = \mathrm{vec}(\Omega)$, and Γ the set of values γ such that $\Omega(\gamma)$ is symmetric and positive definite, for all $(\theta, \gamma) \in \Theta \times \Gamma$, under mild conditions

$$(6) \qquad \begin{pmatrix} g_T(\beta_0) \\ \Delta g_T(\beta_0) \end{pmatrix} \to_d \begin{pmatrix} g \\ \Delta g \end{pmatrix} \sim N\left(\begin{pmatrix} m \\ \mu \end{pmatrix}, \begin{pmatrix} I & \Sigma_{g\theta} \\ \Sigma_{\theta g} & \Sigma_{\theta\theta} \end{pmatrix} \right)$$

so (1) and (2) hold here with $m = \Omega_{ff}^{-\frac{1}{2}} Q_Z c(\beta - \beta_0)$, $\mu = \Omega_{ff}^{-\frac{1}{2}} Q_Z c \in \mathbb{M} = \mathbb{R}^k$, $\Sigma_{g\theta} = \Omega_{ff}^{-\frac{1}{2}} \Omega_{f\beta} \Omega_{ff}^{-\frac{1}{2}}$, $\Sigma_{\theta\theta} = \Omega_{ff}^{-\frac{1}{2}} \Omega_{\beta\beta} \Omega_{ff}^{-\frac{1}{2}}$, and $Q_Z = \mathrm{plim}_{T\to\infty} \frac{1}{T} Z'Z$. Note that for any μ, $m \in \mathcal{M}(\mu) = \{b \cdot \mu : b \in \mathbb{R}\}$ and $m = 0$ when $\beta = \beta_0$. To derive this limit problem, we have imposed very little structure on the data-generating process, and so can easily accommodate heteroscedastic, clustered, or serially correlated data and other features commonly encountered in applied work.

EXAMPLE II—Minimum Distance: A common approach to estimating econometric models is to choose structural parameters to match some vector of sample moments or reduced-form parameter estimates. In minimum distance or moment-matching models, for θ a $p \times 1$ vector of structural parameters and η a $k \times 1$ vector of reduced-form parameters or moments, the model implies that $\eta = f(\theta)$ for some function f. We assume that $f(\theta)$ is continuously differentiable and that $f(\theta)$ and its Jacobian can be calculated either directly or by simulation. Suppose we have an estimator $\hat{\eta}$ for the reduced-form parameter η that, together with an estimator $\hat{\Omega}_\eta$ for the variance of $\hat{\eta}$, satisfies $\hat{\Omega}_\eta^{-\frac{1}{2}}(\hat{\eta} - \eta) \to_d N(0, I)$. Under strong identification asymptotics, $\hat{\eta} - \eta = O_p(\frac{1}{\sqrt{T}})$ and we have the usual asymptotic distribution for the structural parameter estimates $\hat{\theta} = \arg\min_\theta (\hat{\eta} - f(\theta))' \hat{\Omega}_\eta^{-1} (\hat{\eta} - f(\theta))$ and the standard test statistics. If there is limited information about the structural parameters θ, these approximations may be quite poor. One way to model this issue is to take the variance of the reduced-form parameter estimates to be constant, with $\hat{\Omega}_\eta \to_p \Omega_\eta$ for Ω_η non-degenerate, which implies that $\hat{\eta}$ is not consistent for η. Such sequences can often be justified by modeling the variance of the data-generating process as growing with the sample size. In this case, the nonlinearity of $f(\theta)$ will remain important even in large samples, rendering conventional asymptotic approximations unreliable, and the model will be weakly identified in the sense of I. Andrews and Mikusheva (2016b).[3] Let $g_T(\theta) = \hat{\Omega}_\eta^{-\frac{1}{2}}(\hat{\eta} - f(\theta))$, $\Delta g_T(\theta) = \frac{\partial}{\partial \theta'} g_T(\theta) = \hat{\Omega}_\eta^{-\frac{1}{2}} \frac{\partial}{\partial \theta'} f(\theta)$, and $\hat{\gamma} = \mathrm{vec}(\hat{\Omega}_\eta)$.

[3] Note that this is distinct from the type of weak identification for minimum distance models considered in Magnusson (2010) and Magnusson and Mavroeidis (2010), which relates to asymptotically non-negligible variability in the Jacobian of a distance function. For an application of the results of this paper to the model considered in Magnusson and Mavroeidis (2010), see Section F of the Supplemental Material.

For $\gamma = \text{vec}(\Omega_\eta)$ and Γ again the set of γ values corresponding to symmetric positive definite matrices, we have that under $(\theta, \gamma) \in \Theta \times \Gamma$, $\hat{\gamma} \to_p \gamma$ and

$$\begin{pmatrix} g_T(\theta_0) \\ \text{vec}(\Delta g_T(\theta_0)) \end{pmatrix} \to_d \begin{pmatrix} g \\ \text{vec}(\Delta g) \end{pmatrix} \sim N\left(\begin{pmatrix} m \\ \text{vec}(\mu) \end{pmatrix}, \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}\right), \tag{7}$$

where $m \in \mathcal{M} = \{\Omega_\eta^{-\frac{1}{2}}(f(\theta) - f(\theta_0)) : \theta \in \Theta\}$ and $\mu = \Omega_\eta^{-\frac{1}{2}} \frac{\partial}{\partial \theta'} f(\theta_0)$ (see the Supplemental Material for details).

As these examples highlight, limit problems of the form (2) arise in a wide variety of econometric models with weak identification. In the Supplemental Material, we show that GMM models that are weakly identified in the sense of Stock and Wright (2000) generate limit problems of the form (2), and Example I could be viewed as a special case of this result.[4] As Example II illustrates, however, the limit problem (2) is more general. The Supplemental Material provides another non-GMM example, considering a weakly identified generalized minimum distance model studied by Magnusson and Mavroeidis (2010).[5]

Since the limit problem (2) appears in a wide range of weakly identified contexts, for the next several sections we focus on tests in this limit problem. Similarly to Müller (2011), we consider the problem of testing $H_0 : m = 0$, $\mu \in \mathbb{M}$ against $H_1 : m \in \mathcal{M}(\mu) \setminus \{0\}$, $\mu \in \mathbb{M}$ with the limiting random variables $(g, \Delta g, \gamma)$ observed and seek to derive tests with good properties. In the Supplement, we argue that, under mild assumptions, results for the limit problem (2) can be viewed as asymptotic results along sequences of models satisfying (1).

3. PIVOTAL STATISTICS UNDER WEAK IDENTIFICATION

As noted in the Introduction, under weak identification many commonly used test statistics are no longer asymptotically pivotal under the null. To ad-

[4]Whether the set $\mathcal{M}(\mu)$ imposes meaningful restrictions on m will, however, depend on the particular GMM model under consideration. If $\mathcal{M}(\mu)$ imposes no restriction on m, $\mathcal{M}(\mu) = \mathbb{R}^k$ for all $\mu \in \mathbb{M}$, the plug-in approach developed below will not in general be appealing, as the statistic D tells us nothing about how to weight the S and K statistics. Examples of this type include the nonlinear Euler Equation and Quantile IV models considered in I. Andrews and Mikusheva (2016a). One could address these examples by extending the approach of this paper to use the conditioning statistic of I. Andrews and Mikusheva (2016a), but such an extension is beyond the scope of this paper.

[5]Other examples may be found in Guggenberger and Smith (2005, proofs for Theorems 4 and 6), who showed that such convergence also holds in weakly identified Generalized Empirical Likelihood (GEL) models with independent data, both with and without strongly identified nuisance parameters. Guggenberger, Ramalho, and Smith (2012, proofs for Theorems 3.2 and 4.2) extended these results to time-series GEL applications, further highlighting the relevance of the limit problem (2).

dress this issue, much of the literature on identification-robust testing has focused on deriving statistics that are asymptotically pivotal or conditionally pivotal even when identification is weak. Many of the statistics proposed in this literature can be written as functions of the S statistic of Stock and Wright (2000) and the K and D statistics of Kleibergen (2005), or their analogs in non-GMM settings. In this section, we define these statistics, which will play a central role in the remainder of the paper, and develop some results concerning their properties.

When testing $H_0 : m = 0$, $\mu \in \mathbb{M}$ in (2), a natural statistic is

$$S = g'g \sim \chi^2_k(m'm). \tag{8}$$

Under the null S is χ^2 distributed with k degrees of freedom, while under the alternative it is non-central χ^2 distributed with non-centrality parameter $m'm = \|m\|^2$. Statistics asymptotically equivalent to (8) for appropriately defined g_T have been suggested in a number of contexts by a wide range of papers, including Anderson and Rubin (1949) for linear IV, Stock and Wright (2000) for GMM, Magnusson and Mavroeidis (2010) for minimum distance models, and Ramalho and Smith (2004), Otsu (2006), Guggenberger and Smith (2005, 2008), and Guggenberger, Ramalho, and Smith (2012) for generalized empirical likelihood (GEL) models.

While S is a natural statistic for testing $H_0 : m = 0$, $\mu \in \mathbb{M}$, tests based on this statistic are inefficient under strong identification in over-identified models—see Kleibergen (2005). To overcome this problem, Moreira (2001) and Kleibergen (2002) proposed a weak identification-robust score statistic for linear IV models which is efficient under strong identification, and Kleibergen (2005) generalized this statistic to GMM. Following Kleibergen (2005), define D as the $k \times p$ matrix such that

$$\text{vec}(D) = \text{vec}(\Delta g) - \Sigma_{\theta g} g \tag{9}$$

and note that

$$\begin{pmatrix} g \\ \text{vec}(D) \end{pmatrix} \sim N\left(\begin{pmatrix} m \\ \text{vec}(\mu_D) \end{pmatrix}, \begin{pmatrix} I & 0 \\ 0 & \Sigma_D \end{pmatrix} \right),$$

where $\text{vec}(\mu_D) = \text{vec}(\mu) - \Sigma_{\theta g} m$, $\mu_D \in \mathbb{M}_D$, $\Sigma_D = \Sigma_{\theta\theta} - \Sigma_{\theta g}\Sigma_{g\theta}$, and $m \in \mathcal{M}_D(\mu_D)$,

$$\mathcal{M}_D(\mu_D) = \{m : m \in \mathcal{M}(\mu) \text{ for } \text{vec}(\mu) = \text{vec}(\mu_D) + \Sigma_{\theta g} m\}.$$

$\mathcal{M}_D$ plays a role similar to $\mathcal{M}$, defining the set of values m consistent with a given mean μ_D for D. The matrix D can be interpreted as the part of Δg that is uncorrelated with g which, since D and g are jointly normal, implies that D and g are independent. In many models, D is informative about identification:

in linear IV (Example I), for instance, D is a transformation of a particular first-stage parameter estimate.

Kleibergen defined the K statistic as

$$K = g'D\big(D'D\big)^{-1}D'g = g'P_D g. \tag{10}$$

Under the null K is independent of D, and has a χ^2_p distribution. Kleibergen (2005) showed that in GMM, his K test is a score test based on the continuously updating GMM objective function, and subsequent work has developed related statistics in a number of other settings, all of which yield the K statistic (10) in the appropriately defined limit problem. In particular, Magnusson and Mavroeidis (2010) proposed such a statistic for weakly identified generalized minimum distance models, while Ramalho and Smith (2004), Guggenberger and Smith (2005, 2008), and Guggenberger, Ramalho, and Smith (2012) discussed analogs of K for GEL models.

Kleibergen (2005) defined J as the difference between the S and K statistics

$$J = S - K = g'\big(I - D\big(D'D\big)^{-1}D\big)g = g'(I - P_D)g$$

and noted that under the null J is χ^2_{k-p} distributed and is independent of (K, D). Moreira (2003) considered the problem of testing hypotheses on the parameter β in weak IV (Example I) when the instruments Z are fixed and the errors V are normal and homoscedastic with known variance. Moreira derived a conditional likelihood ratio statistic which, for $p = 1$ and $r(D) = D'\Sigma_D^{-1}D$, is

$$\frac{1}{2}\Big(K + J - r(D) + \sqrt{\big(K + J + r(D)\big)^2 - 4J\cdot r(D)}\Big). \tag{11}$$

Under the null the CLR statistic has distribution

$$\frac{1}{2}\Big(\chi^2_p + \chi^2_{k-p} - r(d) + \sqrt{\big(\chi^2_p + \chi^2_{k-p} + r(d)\big)^2 - 4\chi^2_{k-p}\cdot r(d)}\Big) \tag{12}$$

conditional on $D = d$, where χ^2_p and χ^2_{k-p} are independent χ^2 random variables with p and $k - p$ degrees of freedom, respectively. The size α CLR test then rejects when the CLR statistic (11) exceeds $q_\alpha(r(D))$, the $1 - \alpha$ quantile of (12) for $d = D$.

Given this definition, it is natural to consider the class of quasi-CLR (QCLR) tests obtained by using other functions $r : D \to \mathbb{R} \cup \{\infty\}$, where for $r(D) = \infty$ we define the QCLR statistic (11), denoted by $QCLR_r$, to equal K. This class nests the quasi-CLR tests of Kleibergen (2005), Smith (2007), and Guggenberger, Ramalho, and Smith (2012).

For the remainder of the paper, we will focus on the class of tests that can be written as functions of the S, K, and D statistics. While, as the discussion above suggests, this class includes most of the identification-robust procedures proposed in the literature to date, it does rule out some robust tests. In particular, D. Andrews and Cheng (2012, 2013) and D. Andrews and Guggenberger (2014) derived identification-robust tests that cannot in general be written as functions of (S, K, D) and so fall outside the class studied in this paper. Likewise, I. Andrews and Mikusheva (2016a) and I. Andrews and Mikusheva (2016b) considered tests which in general fall outside the class considered here. Further, except in special cases, weighted average power optimal tests based on $(g, \Delta g)$, and in particular the tests proposed by MM for linear IV, will depend on g through more than just S, K, and D and so fall outside this class.

3.1. *Distribution of J and K*

Since the J and K statistics will play a central role in the remainder of the analysis, we discuss their respective properties in the model (2). Note that conditional on $D = d$ for d full rank, the K and J statistics are independent with distribution $K|D = d \sim \chi^2_p(\tau_K(d, m))$ and $J|D = d \sim \chi^2_{k-p}(\tau_J(d, m))$, where

$$\tau_K(D, m) = m' P_D m, \quad \tau_J(D, m) = m'(I - P_D)m. \tag{13}$$

The K statistic picks out a particular (random) direction corresponding to the span of D and restricts attention to deviations from $m = 0$ along this direction. In contrast to the K statistic, the S statistic treats all deviations from $m = 0$ equally and its power depends only on $\|m\|$, which may be quite appealing in cases where $\mathcal{M}_D(\mu_D)$ imposes few restrictions on the possible values of m. To give a sense of the properties of the K statistic, we return to Examples I and II introduced above.

EXAMPLE II—Minimum Distance (Continued): We established in (7) that Δg is non-random, so $D = \Delta g = \mu = \mu_D$. To simplify the exposition, assume for this section that $\Omega_\eta = I$. Since $\mathcal{M} = \{(f(\theta) - f(\theta_0)) : \theta \in \Theta\}$, we have that under alternative θ, the non-centrality parameters in the J and K statistics are

$$\begin{aligned}\big(\tau_J(\theta), \tau_K(\theta)\big) = \Big(&\big(f(\theta) - f(\theta_0)\big)'(I - P_\mu)\big(f(\theta) - f(\theta_0)\big), \\ &\big(f(\theta) - f(\theta_0)\big)' P_\mu \big(f(\theta) - f(\theta_0)\big)\Big).\end{aligned}$$

Since $\mu = \frac{\partial}{\partial\theta'} f(\theta_0)$, this means that under alternative θ, the non-centrality parameter τ_K is the squared length of $f(\theta) - f(\theta_0)$ projected onto the model's tangent space at the null parameter value, while τ_J is the squared length of the residual from this projection. If $f(\theta)$ is linear so $f(\theta) = \frac{\partial}{\partial\theta'} f(\theta_0)(\theta - \theta_0)$ and $\mathcal{M} = \{\frac{\partial}{\partial\theta'} f(\theta_0) \cdot b : b \in \mathbb{R}^p\}$, $\tau_J \equiv 0$ and the K test ϕ_K will be uniformly most powerful in the class of tests based on (S, K, D). As argued in I. Andrews and Mikusheva (2016b), under conventional (strong-identification) asymptotics, minimum distance models are approximately linear, confirming the desirable properties of the K statistic in this case. Under weak identification, however, nonlinearity of $f(\theta)$ may remain important even asymptotically. To take an extreme case, if there is some $\theta \in \Theta$ such that $\|f(\theta)-f(\theta_0)\|>0$ and $\frac{\partial}{\partial\theta}f(\theta_0)'(f(\theta)-f(\theta_0))=0$, the K statistic will not help in detecting such an alternative and the optimal test against θ based on (J,K) depends on J alone.

EXAMPLE I—Weak IV (Continued): In the limit problem (6), Δg is random and may be correlated with g, so $D\neq\Delta g$ and $\mu_D=\mu-\Sigma_{\theta g}m$. Since $m=\mu(\beta-\beta_0)$,

$$\mu_D=\mu-\Sigma_{\theta g}\mu(\beta-\beta_0)=\big(I-\Sigma_{\theta g}(\beta-\beta_0)\big)\mu.$$

Note that if μ is proportional to an eigenvector of $\Sigma_{\theta g}$ corresponding to a nonzero eigenvalue λ, then for $(\beta-\beta_0)=\lambda^{-1}$ we have that $\mu_D=0$. Hence for some $(\Sigma_{\theta g},\mu)$ combinations, while μ may be quite large relative to both $\Sigma_{\theta g}$ and $\Sigma_{\theta\theta}$, there will be some alternatives β under which $\mu_D=0$. When this occurs, the direction of the vector D bears no relation to the direction of m or μ and the K statistic picks a direction entirely at random and so loses much of its appeal. The well-known non-monotonicity of the power function for tests based on K is a consequence of this fact. If, as in the homoscedastic model considered by AMS, Ω as defined in (5) has Kronecker product structure $\Omega=A\otimes B$ for a 2×2 matrix A and a $k\times k$ matrix B, then there is a value $\beta=\beta_{AR}$ defined by AMS where $\mu_D=0$ regardless of the true value μ.

The case where Ω has Kronecker product structure is extreme in that $\mu_D=0$ at β_{AR} regardless of the true value μ. However, tests based on the K statistic face other challenges in the non-Kronecker case. In particular, in the Kronecker product case $\mu_D\propto\mu$ and so, as long as $\mu_D\neq 0$, the mean of D has the correct direction, while in contrast, $\mu_D\not\propto\mu$ in the general (non-Kronecker) case. An extreme version of this issue arises if there is some value β^* such that $(I-\Sigma_{\theta g}(\beta^*-\beta_0))\mu\neq 0$ but $\mu'(I-\Sigma_{\theta g}(\beta^*-\beta_0))\mu=0$. For this value of β^*, we have that $\mu_D\neq 0$ but $\mu_D'm=0$, and hence the K statistic tends to focus on directions that yield low power against alternative β^*. Likewise, if we try to assess the reliability of the K statistic by measuring the norm of D, we risk being led astray in this case.

To summarize, tests based on K have good power when the direction of D is similar to that of m, but may have poor power otherwise. By contrast, the S test (8) has power that depends only on $\|m\|$ and thus does not suffer from the spurious loss of power that can affect tests based on K. The question in constructing tests based on (S,K,D) (or equivalently (J,K,D)) is thus how to use the information contained in D to combine the S and K statistics to retain the advantages of each while ameliorating their deficiencies.

4. OPTIMAL TESTS IN A CONDITIONAL PROBLEM

After restricting attention to tests that depend on the data only through (S,K,D), we are interested in constructing powerful tests for the null H_0 :

$m = 0$, $\mu \in \mathbb{M}$ against the alternative $H_1 : m \in \mathcal{M}(\mu) \setminus \{0\}$, $\mu \in \mathbb{M}$. As a first step, we consider the subproblem that arises after we condition on the realized value of D. Conditional on the event $D = d$ (for d full rank), J and K are independent and distributed $\chi^2_{k-p}(\tau_J(d, m))$ and $\chi^2_p(\tau_K(d, m))$, respectively, for τ_J and τ_K as defined in (13). Once we condition on $D = d$, our null hypothesis $H_0 : m = 0$, $\mu \in \mathbb{M}$ can be rewritten as $H_0 : \tau_J = \tau_K = 0$.

A first task is to characterize the set of possible values for the non-centrality parameters (τ_J, τ_K) under the alternative H_1. Let $\mathbb{M}_D(d)$ denote the set of values $\mu_D \in \mathbb{M}_D$ such that d is in the support of D. If Σ_D is full rank, then $\mathbb{M}_D(d) = \mathbb{M}_D$, since the support of D is the same for all $\mu_D \in \mathbb{M}_D$, but if Σ_D is reduced rank (e.g., $\Sigma_D = 0$), then we may have $\mathbb{M}_D(d) \subset \mathbb{M}_D$. Letting $\widetilde{\mathcal{M}}(d) = \bigcup_{\mu_D \in \mathbb{M}_D(d)} \mathcal{M}(\mu_D)$, we see that conditional on $D = d$, m may take any value in $\widetilde{\mathcal{M}}(d)$ and still be consistent with both $m \in \mathcal{M}(\mu_D)$ and d lying in the support of D. Hence, the non-centrality parameters (τ_J, τ_K) may take any value in the set

$$\mathcal{T}(d) = \bigcup_{m \in \widetilde{\mathcal{M}}(d)} \bigl(\tau_J(d, m), \tau_K(d, m)\bigr). \tag{14}$$

Conditional on $D = d$, our problem becomes one of testing $H_0 : \tau_J = \tau_K = 0$ against the alternative $H_1 : (\tau_J, \tau_K) \in \mathcal{T}(d) \setminus \{0\}$ based on observing $(J, K) \sim (\chi^2_{k-p}(\tau_J), \chi^2_p(\tau_K))$. Note that in Example I (linear IV), we have that $\mathcal{T}(d) = \mathbb{R}^2_+$ for all d provided $\mathbb{M} = \mathbb{R}^k$ and Σ_D is full rank.

If either τ_J or τ_K is known to be zero, then there is a uniformly most powerful test conditional on $D = d$. In particular, if we know that $\tau_K = 0$ ($\mathcal{T}(d) = A \times \{0\}$ for $A \neq \{0\}$), then the test

$$\phi_J = 1\bigl\{J > \chi^2_{k-p,1-\alpha}\bigr\} \tag{15}$$

is a uniformly most powerful level α test of $H_0 : \tau_J = \tau_K = 0$ against $H_1 : (\tau_J, \tau_K) \in \mathcal{T}(d) \setminus \{0\}$.[6] Likewise, if τ_J is known to be zero ($\mathcal{T}(d) = \{0\} \times A$), then the test

$$\phi_K = 1\bigl\{K > \chi^2_{p,1-\alpha}\bigr\} \tag{16}$$

is uniformly most powerful. Unfortunately, when neither τ_K nor τ_J is restricted to be zero, there is not in general a uniformly most powerful test in this problem, and different alternatives (τ_J, τ_K) imply different optimal tests. To proceed, we study the class of tests that are admissible in the conditional problem in the case with $\mathcal{T}(d) = \mathbb{R}^2_+$.

[6]For a non-randomized test ϕ, we denote rejection by $\phi = 1$ and non-rejection by $\phi = 0$.

4.1. *Admissible Tests in the Conditional Problem*

A result from Marden (1982) establishes that the class of admissible tests in the conditional problem has a simple form when $\mathcal{T}(d) = \mathbb{R}^2_+$.

THEOREM 1—Marden (1982): *Let $J \sim \chi^2_{k-p}(\tau_J)$ and $K \sim \chi^2_p(\tau_K)$ be independent and let ϕ be a test of $H_0 : \tau_J = \tau_K = 0$ against $H_1 : (\tau_J, \tau_K) \in \mathcal{T}(d) \setminus \{0\}$ for $\mathcal{T}(d) = \mathbb{R}^2_+$. ϕ is admissible in the class of tests which depend on (J, K) if and only if it is almost surely equal to $1\{(\sqrt{J}, \sqrt{K}) \notin C\}$ for some set C satisfying*

1. *C is closed and convex,*
2. *C is monotone decreasing, that is, $x \in C$ and $y_i \leq x_i$ $\forall i$ implies $y \in C$.*

Thus, a test ϕ is admissible in the conditional problem only if its acceptance region in $(\sqrt{J}, \sqrt{K})$ space is almost everywhere equal to a closed, convex, monotone decreasing set. It is important to note that Theorem 1 concerns only admissibility in the problem where we have conditioned on $D = d$. Admissibility in this conditional problem for all values d is not sufficient for admissibility as a test of $H_0 : m = 0$, $\mu \in \mathbb{M}$ against $H_1 : m = \mathcal{M}(\mu) \setminus \{0\}$, $\mu \in \mathbb{M}$ in the original problem. However, the set of tests which are admissible in the conditional problem for almost all d form an essentially complete class: that is, for any test ϕ, we can find a test $\tilde{\phi}$ which has weakly lower size and higher power than ϕ and is admissible in the conditional problem for almost every d (with respect to the distribution of D, F_D).

Theorem 1 implies that the tests ϕ_J and ϕ_K defined in (15) and (16) are admissible in the conditional problem for all non-singular d with $\mathcal{T}(d) = \mathbb{R}^2_+$ and all significance levels. The test

$$\phi_S = 1\{S > \chi^2_{k,1-\alpha}\} \tag{17}$$

is likewise admissible for all d with $\mathcal{T}(d) = \mathbb{R}^2_+$. For all functions $r : D \to \mathbb{R}_+ \cup \{\infty\}$, the level α QCLR test based on the statistic $QCLR_r$ defined in (11) is admissible for all d with $\mathcal{T}(d) = \mathbb{R}^2_+$ as well: this follows immediately from Theorem 1 together with Theorem 4 below. More broadly, we can see that for any functions $A(j, k, d)$ and $c(d)$ such that A has closed, convex, and monotone decreasing lower sets in $(\sqrt{j}, \sqrt{k})$ space for all d, the test that rejects when $A(J, K, D) > c(D)$ will be admissible in the conditional problem for all d with $\mathcal{T}(d) = \mathbb{R}^2_+$. Thus, the class of tests which are potentially admissible in the conditional problem is still extremely large, so we consider two additional optimality criteria in the conditional problem: local power and weighted average power.

4.2. *Optimality of Linear Tests in the Conditional Problem*

Tests which reject when a linear combination of the J and K statistics is large are both locally most powerful against sequences of alternatives approaching

$(\tau_J, \tau_K) = 0$ linearly and weighted average power maximizing for scaled χ^2 weight functions.

THEOREM 2—Monti and Sen (1976), Koziol and Perlman (1978): *Let* $J \sim \chi^2_{k-p}(\tau_J)$ *and* $K \sim \chi^2_p(\tau_K)$ *be independent and let* Φ_α *denote the class of size* α *tests of* $H_0 : \tau_J = \tau_K = 0$ *against* $H_1 : (\tau_J, \tau_K) \in \mathcal{T}(d) \setminus \{0\}$ *for* $\mathcal{T}(d) = \mathbb{R}^2_+$. *Fix a constant* $a \geq 0$, *and let* $\phi_a = 1\{K + a \cdot J > c_\alpha(a)\}$ *for* $c_\alpha(a)$ *the* $1 - \alpha$ *quantile of a* $\chi^2_p + a \cdot \chi^2_{k-p}$ *distribution*.

1. *Let* $(\tau_J, \tau_K) = \lambda \cdot (a\frac{k-p}{p}, 1)$. *For any test* $\phi \in \Phi_\alpha$, *there exists* $\bar{\lambda} > 0$ *such that if* $0 < \lambda < \bar{\lambda}$,

$$E_{(\tau_J,\tau_K)}[\phi] \leq E_{(\tau_J,\tau_K)}[\phi_a].$$

2. *Let* $F_{t_J,t_K}(\tau_J, \tau_K)$ *be the distribution function for* $(\tau_J, \tau_K) \sim (t_J \cdot \chi^2_{k-p}, t_K \cdot \chi^2_p)$. *For any* (t_J, t_K) *with* $\frac{t_J}{t_K}\frac{t_K+1}{t_J+1} = a$, *the test* ϕ_a *solves the weighted average power maximization problem*

$$\phi_a \in \arg\max_{\phi \in \Phi_a} \int E_{(\tau_J,\tau_K)}[\phi_{a(D)}]\, dF_{t_J,t_K}(\tau_J, \tau_K).$$

Theorem 2 follows immediately from results in Monti and Sen (1976) and Koziol and Perlman (1978) on the optimal combination of independent noncentral χ^2 statistics. Theorem 2(1) shows that the test based on $K + a \cdot J$ is locally most powerful against sequences of alternatives with $\tau_J/\tau_K = a\frac{k-p}{p}$. Theorem 2(2) establishes that ϕ_a maximizes weighted average power in the conditional problem for a continuum of different weight functions corresponding to scaled χ^2 distributions.

EXAMPLE I—Weak IV (Continued): In the linear IV model, $\tau_K = (\beta - \beta_0)^2(\mu' d)^2/d'd$, while $\tau_J = (\beta - \beta_0)^2\mu'\mu - \tau_K$. Thus, one can show that for φ the angle between μ and d, $\tau_K = (\beta - \beta_0)^2\mu'\mu\cos^2(\varphi)$, $\tau_J = (\beta - \beta_0)^2\mu'\mu\sin^2(\varphi)$, and $\tau_J/\tau_K = \tan^2(\varphi)$. Hence, if φ were known, the locally most powerful test in the conditional problem would take $a(d) = \frac{1}{k-1}\tan^2(\varphi)$. Consequently, if we knew that the direction of d were similar to that of μ, the locally most powerful test would heavily weight K, while if we thought the directions were different, we would prefer to put more weight on J. The form of the weighted average power optimal test is less transparent, but again shows that when we think the direction of μ and d is similar, we want to put most weight on K, while when their directions differ, we want to put more weight on J.

Theorem 2 shows that, among the large class of tests which are admissible in the conditional problem, tests based on linear combinations of the J

and K statistics have particularly good properties. In addition to satisfying the conditional admissibility requirement given by Theorem 1, they maximize local power against sequences of alternatives with $\frac{\tau_J}{(k-p)}\frac{p}{\tau_K}=a$. Indeed, one can show that the power of any size α test ϕ in the conditional problem is (weakly) exceeded by the power envelope for the class of size α linear combination tests once $\tau_J+\tau_K$ is sufficiently small, regardless of the direction of (τ_J,τ_K). Against non-local deviations, on the other hand, Theorem 2(2) shows that the linear combination test with weight a on J maximizes weighted average power for a continuum of scaled χ^2 weights with $\frac{t_J}{t_K}\frac{t_K+1}{t_J+1}=a$. As noted by Koziol and Perlman (1978), taken together these results show that tests based on linear combinations of the J and K statistics have good power against a wide range of alternatives in the conditional problem. In the next section, we define the class of tests that yield linear combination tests conditional on $D=d$ and show that CLR and Quasi-CLR tests are members of this class.

5. CONDITIONAL LINEAR COMBINATION TESTS

If we restrict attention to tests which yield level α linear combination tests conditional on $D=d$, we obtain the class of conditional linear combination tests. For a weight function $a:\mathcal{D}\to[0,1]$, the corresponding conditional linear combination test, $\phi_{a(D)}$, rejects when a convex combination of the S and K statistics weighted by $a(D)$ exceeds a conditional critical value[7]:

$$\begin{aligned}\phi_{a(D)}&=1\big\{\big(1-a(D)\big)\cdot K+a(D)\cdot S>c_\alpha\big(a(D)\big)\big\}\\&=1\big\{K+a(D)\cdot J>c_\alpha\big(a(D)\big)\big\}.\end{aligned}\tag{18}$$

We take the conditional critical value $c_\alpha(a)$ to be the $1-\alpha$ quantile of a $\chi^2_p+a\cdot\chi^2_{k-p}$ distribution. This choice ensures that $\phi_{a(D)}$ will be conditionally similar, and thus similar, for any choice of $a(D)$. Stated formally, we have the following:

THEOREM 3: *For any weight function* $a:\mathcal{D}\to[0,1]$, *the test* $\phi_{a(D)}$ *defined in* (18) *is conditionally similar with* $E_{m=0,\mu_D}[\phi_{a(D)}|D]=\alpha$ *almost surely for all* $\mu_D\in\mathbb{M}_D$. *Hence*, $E_{m,\mu_D}[\phi_{a(D)}]=\alpha$ *for all* $(m,\mu_D)\in H_0$ *and* $\phi_{a(D)}$ *is a similar test*.

While we could construct a family of CLC tests based on some conditional critical value function other than $c_\alpha(a)$ that does not impose conditional similarity, restricting attention to conditionally similar tests is a simple way to ensure correct size regardless of our choice of $a(D)$, and is equivalent to similarity

[7]Note that in defining conditional linear combination tests here, we restrict to $a\in[0,1]$, so that the S and K tests are the extremes in the class. The extension to $a\in\mathbb{R}_+$ is straightforward.

if $\mathbb{M}_D$ contains an open set, as will typically be the case (e.g., in Example I). Restricting attention to similar CLC tests has the further advantage that all such tests are unbiased, in that their power against any alternative is at least equal to α:

LEMMA 1: *For any weight function* $a : \mathcal{D} \to [0, 1]$, *the test* $\phi_{a(D)}$ *defined in* (18) *is unbiased both conditional on* D *and unconditionally, in the sense that for any* m, $E_{m,\mu_D}[\phi_{a(D)}|D] \geq \alpha$ *almost surely, and* $E_{m,\mu_D}[\phi_{a(D)}] \geq \alpha$.

Interestingly, the class of QCLR tests is precisely the same as the class of CLC tests. Formally, for any function $r : \mathcal{D} \to \mathbb{R}_+ \cup \{\infty\}$, define the quasi-CLR statistic $QCLR_r$ as in (11) and let $q_\alpha(r(d))$ be the $1 - \alpha$ quantile of (12). Then we have the following:

THEOREM 4: *For any function* $r : \mathcal{D} \to \mathbb{R}_+ \cup \{\infty\}$, *if we take*

$$\phi_{QCLR_r} = 1\{QCLR_r > q_\alpha\big(r(D)\big)\},$$

then for $\tilde{a}(D) = \frac{q_\alpha(r(D))}{q_\alpha(r(D))+r(D)}$ *we have* $\phi_{QCLR_r} \equiv \phi_{\tilde{a}(D)}$. *Conversely, for any* $a : \mathcal{D} \to [0, 1]$, *there exists an* $\tilde{r} : \mathcal{D} \to \mathbb{R}_+ \cup \{\infty\}$ *such that* $\phi_{a(D)} \equiv \phi_{QCLR_{\tilde{r}}}$. *Hence, the class of CLC tests for* $a : \mathcal{D} \to [0, 1]$ *is precisely the same as the class of QCLR tests for* $r : \mathcal{D} \to \mathbb{R}_+ \cup \{\infty\}$.

Theorem 4 shows that the QCLR test ϕ_{QCLR_r} is a conditional linear combination test with weight function $a(D) = \frac{q_\alpha(r(D))}{q_\alpha(r(D))+r(D)}$. In particular, this result establishes that the CLR test of Moreira (2003) for linear IV with a single endogenous regressor is a CLC test. In the remainder of the paper, our exposition focuses on CLC tests, but by Theorem 4, all of our results apply to QCLR tests as well. An immediate corollary of this result, together with Lemma 1, is that all QCLR tests are unbiased. This is, to the best of our knowledge, a new result even for the CLR test.[8]

COROLLARY 1: *For any function* $r : \mathcal{D} \to \mathbb{R}_+ \cup \{\infty\}$, *the Quasi-CLR test* ϕ_{QCLR_r} *is unbiased both conditional on* D *and unconditionally, in the sense that for any* m, $E_{m,\mu_D}[\phi_{QCLR_r}|D] \geq \alpha$ *almost surely and* $E_{m,\mu_D}[\phi_{QCLR_r}] \geq \alpha$.

It is worth highlighting a subtlety in Theorem 4. While $\phi_{QCLR_r} \equiv \phi_{\tilde{a}(D)}$, it is *not* the case that $QCLR_r = K + \tilde{a}(D) \cdot J$, nor that $QCLR_r = f(K + \tilde{a}(D) \cdot J)$ for any function f. Indeed, for fixed D, the level sets of $K + \tilde{a}(D) \cdot J$ in (J, K) space will be linear and parallel. By contrast, while the proof of Theorem 4 shows that the level sets of $QCLR_r$ are linear, they are not parallel. Thus, while for

[8] AMS showed that the CLR test satisfies a necessary condition for unbiasedness, but not that it is unbiased.

any given α and d we can choose $\tilde{a}$ so that $QCLR_r$ exceeds its $1-\alpha$ conditional quantile if and only if $K + \tilde{a}(d) \cdot J$ exceeds its own $1-\alpha$ conditional quantile, the QCLR and CLC test statistics are not equivalent in some broader sense.

6. OPTIMAL CLC TESTS

For any weight function $a : \mathcal{D} \to [0, 1]$, we can define a CLC test $\phi_{a(D)}$ for H_0 against H_1 using (18). While any such test controls size by Theorem 3, the class of such CLC tests is large and we would like a systematic way to pick weight functions a yielding tests with good power properties.

A natural optimality criterion, after restricting attention to CLC tests, is minimax regret. To define a minimax regret CLC test, for any $(m, \mu_D) \in H_1$ define $\beta^*_{m,\mu_D} = \sup_{a\in\mathcal{A}} E_{m,\mu_D}[\phi_{a(D)}]$ for $\mathcal{A}$ the class of measurable functions $a : \mathcal{D} \to [0, 1]$. β^*_{m,μ_D} gives the highest attainable power against alternative (m, μ_D) in the class of CLC tests and, as we vary (m, μ_D), defines the power envelope for this class. For a given $a \in \mathcal{A}$, we can then define the regret associated with $\phi_{a(D)}$ against alternative (m, μ_D) as $\beta^*_{m,\mu_D} - E_{m,\mu_D}[\phi_{a(D)}]$, which is the amount by which the power of the test $\phi_{a(D)}$ falls short of the highest power we might have attained against this alternative by choosing some other CLC test. We can then define the maximum regret for a test $\phi_{a(D)}$ as $\sup_{(m,\mu_D)\in H_1}(\beta^*_{m,\mu_D} - E_{m,\mu_D}[\phi_{a(D)}])$, which is the largest amount by which the power function of $\phi_{a(D)}$ falls short of the power envelope for the class of CLC tests. A minimax regret choice of $a \in \mathcal{A}$ is

$$a_{\mathrm{MMR}} \in \arg\min_{a\in\mathcal{A}} \sup_{(m,\mu_D)\in H_1} \left(\beta^*_{m,\mu_D} - E_{m,\mu_D}[\phi_{a(D)}]\right).$$

As an optimality criterion, this is an intuitive choice: having already restricted attention to the class of CLC tests, focusing on MMR tests minimizes the maximal extent to which the test we choose could under-perform relative to other CLC tests.

EXAMPLE II—Minimum Distance (Continued): Calculating the MMR test in Example II is straightforward. In particular, $D = \mu$ is non-random, so rather than picking a function from $\mathcal{D}$ to $[0, 1]$, we are simply picking a number a in $[0, 1]$. Moreover, we know that in this example $m = m(\theta) = \Omega_\eta^{-\frac{1}{2}}(f(\theta) - f(\theta_0))$ and $\mu_D = \mu = \Omega_\eta^{-\frac{1}{2}} \frac{\partial}{\partial\theta} f(\theta_0)$, so the maximum attainable power against alternative θ is simply $\beta^*_\theta = \sup_{a\in[0,1]} E_{m(\theta),\mu}[\phi_a]$ which we can calculate for any value θ. To solve for the MMR test ϕ_{MMR}, we need only calculate $a_{\mathrm{MMR}} = \arg\min_{a\in[0,1]} \sup_{\theta\in\Theta}(\beta^*_\theta - E_{m(\theta),\mu}[\phi_a])$.

6.1. *Plug-in Minimax Regret Tests*

While finding the MMR test is straightforward in Example II, Example I is less tractable in this respect. In this example, D is random, so solving for

ϕ_{MMR} requires that we optimize over the class $\mathcal{A}$ of functions. In most cases, finding even an approximate solution to this optimization problem is extremely computationally costly, rendering ϕ_{MMR} unattractive in many applications. To overcome this difficulty, we suggest a computationally tractable class of plug-in tests.

There are two aspects of Example II which make calculating ϕ_{MMR} straightforward. First, rather than optimizing over the space of functions $\mathcal{A}$, we need only optimize over numbers in $[0, 1]$. Second, $\mu = \mu_D$ is known, so in solving the minimax problem we need only search over $\theta \in \Theta$ rather than over some potentially higher dimensional space of values for $(m, \mu_D) \in H_1$.

To construct a test for the general case with similarly modest computational requirements, imagine first that μ_D is known. Let us restrict attention to unconditional linear combination tests with $a(D) \equiv a(\mu_D) \in [0, 1]$. The power envelope for this class of *unconditional* linear combination tests is $\beta^u_{m,\mu_D} = \sup_{a\in[0,1]} E_{m,\mu_D}[\phi_a]$. A minimax regret unconditional (MMRU) test $\phi_{\text{MMRU}} = \phi_{a_{\text{MMRU}}(\mu_D)}$ then uses

$$a_{\text{MMRU}}(\mu_D) \in \arg\min_{a\in[0,1]} \sup_{m\in\mathcal{M}_D(\mu_D)} \big(\beta^u_{m,\mu_D} - E_{m,\mu_D}[\phi_a]\big). \tag{19}$$

Just as when we derived ϕ_{MMR} for Example II above, here we need only optimize over $a \in [0, 1]$ and $m \in \mathcal{M}_D(\mu_D)$, rather than over $a \in \mathcal{A}$ and $(m, \mu_D) \in H_1$.[9]

In defining ϕ_{MMRU}, we assumed that μ_D was known, which is unlikely to hold in contexts like Example I where D is random. Note, however, that for any estimator $\hat{\mu}_D$ which depends only on D, $a_{\text{MMRU}}(\hat{\mu}_D)$ can be viewed as a particular weight function $a(D)$ and the *plug-in* minimax regret (PI) test

$$\phi_{\text{PI}} = \phi_{a_{\text{PI}}(D)} = 1\big\{K + a_{\text{MMRU}}(\hat{\mu}_D) \cdot J > c_\alpha\big(a_{\text{MMRU}}(\hat{\mu}_D)\big)\big\} \tag{20}$$

is a CLC test and so controls size by Theorem 3. Moreover, to calculate this test, we need only solve for a_{MMRU} taking the estimate $\hat{\mu}_D$ to be the true value, so this test remains computationally tractable.

It is important to note that ϕ_{PI} is not in general a true MMR test. First, ϕ_{PI} treats the estimated value $\hat{\mu}_D$ as the true value, and hence does not account for any uncertainty in the estimation of μ_D. Second, even taking the value μ_D as given, ϕ_{PI} restricts attention to unconditional linear combination tests, which represent a strict subset of the possible functions $a \in \mathcal{A}$. Despite these potential shortcomings, we find that PI tests perform quite well in simulation, and show in the Supplement that PI tests will be asymptotically optimal under strong identification in our examples.

To use PI tests in a given context, we need only choose the estimator $\hat{\mu}_D$. While the MLE for μ_D based on D, $\hat{\mu}_D = D$, is a natural choice, we may be

[9]When the argmin is non-unique, we select the largest value a belonging to the argmin.

able to do better in many cases. In particular, in weak IV (Example I) with homoscedastic errors, estimation of $\hat{\mu}_D$ is related to a problem of non-centrality parameter estimation, allowing us to use results from that literature.

EXAMPLE I—Weak IV (Continued): Consider the case studied by AMS where $\Omega = A \otimes B$ has Kronecker product structure. Results in AMS show that $(J, K, D'\Sigma_D^{-1}D)$ is a maximal invariant under rotations of the instruments, where $D'\Sigma_D^{-1}D \sim \chi_k^2(\mu_D'\Sigma_D^{-1}\mu_D)$.[10] AMS showed that the distribution of $(J, K, D'\Sigma_D^{-1}D)$ depends on $c = \sqrt{T}\pi_T$ only through the non-centrality parameter $r = \mu_D'\Sigma_D^{-1}\mu_D$.

Note that the MLE $\hat{\mu}_D = D$ for μ_D based on D implies a severely biased estimator for r, $\hat{r} = D'\Sigma_D^{-1}D$, with $E[\hat{r}] = E[D'\Sigma_D^{-1}D] = r + k$. The problem of estimating r relates to the well-studied problem of estimating the non-centrality parameter of a non-central χ^2 distribution, and a number of different estimators have been proposed for this purpose, including $\hat{r}_{\mathrm{MLE}}$, the MLE for r based on $\hat{r}$ (which is not available in closed form), and $\hat{r}_{\mathrm{PP}} = \max\{\hat{r} - k, 0\}$, which is the positive part of the bias corrected estimator $\hat{r} - k$.[11] Both $\hat{r}_{\mathrm{MLE}}$ and $\hat{r}_{\mathrm{PP}}$ are zero for a range of values $\hat{r} > 0$ so we also consider an estimator proposed by Kubokawa, Roberts, and Saleh (1993),

$$\hat{r}_{\mathrm{KRS}} = \hat{r} - k + e^{-\frac{\hat{r}}{2}} \left(\sum_{j=0}^{\infty} \left(-\frac{\hat{r}}{2} \right)^j \frac{1}{j!(k+2j)} \right)^{-1},$$

which is smooth in $\hat{r}$ and greater than zero whenever $\hat{r} > 0$. We show in Section 7 below that estimators $\hat{\mu}_D$ corresponding to all three non-centrality estimators $\hat{r}_{\mathrm{MLE}}$, $\hat{r}_{\mathrm{PP}}$, and $\hat{r}_{\mathrm{KRS}}$ yield PI tests ϕ_{PI} with good power properties, where, for each estimator $\hat{r}_i$, we let $\hat{\mu}_D = D \cdot \sqrt{\hat{r}_i/\hat{r}}$.

7. PERFORMANCE OF PI TESTS IN WEAK IV

In this section, we examine the performance of PI tests in linear IV with weak instruments (Example I). For comparability with the previous literature, we first consider the homoscedastic model studied by AMS. Since data encountered in empirical practice commonly violate this homoscedasticity assumption, we then consider the performance of PI tests in a model calibrated to match the heteroscedastic time-series data used by Yogo in his (2004) study on the effect of weak instruments on estimation of the elasticity of intertemporal substitution.

[10]It suffices to note that $(J, K, D'\Sigma_D^{-1}D)$ is a one-to-one transformation of Q as defined in AMS.

[11]$\hat{r}_{\mathrm{PP}}$ has been shown to dominate $\hat{r}_{\mathrm{MLE}}$ in terms of mean squared error but is itself inadmissible (Saxena and Alam (1982)).

7.1. *Homoscedastic Linear IV*

AMS considered the linear IV model Example I with homoscedastic normal errors and showed that in this case the CLR test of Moreira (2003) is nearly uniformly most powerful in a class of two-sided tests invariant to rotations of the instruments, in the sense that the power function of the CLR test is uniformly close to the power envelope for this class. Müller (2011) then showed that the CLR test is nearly asymptotically uniformly most powerful in the class of invariant two-sided tests that have correct size under (6) with the additional restriction that $\Omega = A \otimes B$ for A and B symmetric positive-definite matrices of dimension 2×2 and $k \times k$, respectively. As Mueller noted, matrices Ω of this form arise naturally only for serially uncorrelated homoscedastic IV models, limiting the applicability of this result. Nonetheless, if our plug-in minimax regret approach is to work well in this benchmark case, it should match the near-optimal performance of the CLR test. In the homoscedastic case, one can show that the plug-in weight functions depend on D only through $\hat{r}$, so in Section D.1.1 of the Supplemental Material, we directly compare the plug-in weights to the CLR weight function.

7.1.1. *Power Simulation Results*

To study the power of the PI tests, we follow the simulation design of AMS and consider a homoscedastic normal model with a known reduced-form covariance matrix. Like AMS, we consider models with five instruments, reduced-form error correlation ρ equal to 0.5 or 0.95, and concentration (identification strength) parameter $\lambda = \mu'\mu$ equal to 5 and 20. To examine the effect of changing the number of instruments, as in AMS we also consider models with two and ten instruments, in each case fixing ρ equal to 0.5 and letting λ equal 5 and 20. The resulting power plots (based on 10,000 simulations) are reported in the Supplemental Material. For brevity, here we report only the maximal power shortfall of each test, measured as the maximal distance from the power functions of the CLR, PI, AR, and K tests to the power envelope for the class consisting of these tests alone. These values can be viewed as a measure of maximum regret relative to this restricted set of tests. As Table I makes clear, the PI tests considered largely match the near-optimal performance of the CLR test. The one exception is the PI test using the badly biased estimator $\hat{r}$ for r, which systematically overweights the K statistic and consequently under-performs relative to the other tests considered. As these results highlight, in one of the only weakly identified contexts where a near-UMP test is known, reasonable implementations of the plug-in testing approach suggested in this paper are near-optimal as well.

7.2. *Linear IV With Unrestricted Covariance Matrix*

The near-optimal performance of PI tests in linear IV models where Ω has Kronecker product structure is promising, but is of limited relevance for em-

TABLE I

MAXIMAL POWER SHORTFALL RELATIVE TO OTHER TESTS CONSIDERED, IN LINEAR IV MODEL WITH HOMOSCEDASTIC ERRORS[a]

	CLR	PI-$\hat{r}$	PI-$\hat{r}_{\text{MLE}}$	PI-$\hat{r}_{\text{PP}}$	PI-$\hat{r}_{\text{KRS}}$	AR	K
$k = 2$	1.18%	1.44%	0.72%	0.72%	0.88%	9.40%	29.96%
$k = 5$	2.14%	5.90%	1.37%	1.07%	2.04%	25.05%	53.71%
$k = 10$	3.51%	13.21%	2.29%	2.18%	4.00%	30.76%	64.62%

[a]For each k (number of instruments), we calculate the pointwise maximal power of the tests studied. For each test, we report the largest margin by which the power of that test falls short of pointwise maximal power. CLR denotes the CLR test of Moreira (2003), while PI-$\hat{r}$, PI-$\hat{r}_{\text{MLE}}$, PI-$\hat{r}_{\text{PP}}$, and PI-$\hat{r}_{\text{KRS}}$ denote the PI tests with weight functions $a_{\text{MMRU}}(\hat{r})$, $a_{\text{MMRU}}(\hat{r}_{\text{MLE}})$, $a_{\text{MMRU}}(\hat{r}_{\text{PP}})$, and $a_{\text{MMRU}}(\hat{r}_{\text{KRS}})$, respectively. AR is the Anderson–Rubin test (equivalent to the S test) and K is Kleibergen's (2002) K test.

pirical work. Economic data frequently exhibit heteroscedasticity, serial dependence, clustering, and other features that render Kronecker structure for Ω implausible. It is natural to ask whether PI tests continue to have good power properties in this more general case. As an alternative CLC test, we consider Kleibergen's (2005) quasi-CLR test, which takes $r(D) = D'\Sigma_D^{-1}D$ and can be viewed as a heteroscedasticity and autocorrelation-robust version of the CLR test, as well as the K and Anderson–Rubin (S) tests.

We also consider the conditional QLR test of I. Andrews and Mikusheva (2016a) and the MM1-SU and MM2-SU tests of MM. The conditional QLR test is a direct generalization of the CLR test of Moreira (2003) to the non-homoscedastic case, and rejects when a quasi-likelihood ratio statistic based on the continuously updating GMM objective exceeds its conditional critical value given D. The MM tests, on the other hand, maximize weighted average power, for weights which depend on the covariance matrix Σ, over a class of similar tests satisfying a sufficient condition for local unbiasedness (their SU tests). We show in the Supplemental Material that all conditional linear combination tests are SU tests, and thus that the MM1-SU and MM2-SU tests have, by construction, weighted average power at least as great as all CLC tests under their respective weight functions. MM presented extensive simulation results which show that these tests perform very well in models where Ω has Kronecker structure, as well as in examples with non-Kronecker structure.[12]

There are a multitude of ways in which Ω may depart from Kronecker structure, and it is far from clear ex ante how the power of the tests we consider may be expected to compare under different departures. To assess the

[12]For the MM tests, we follow Moreira and Moreira (2015) and set the tuning parameters σ and ς to one-tenth of the sample size in each case (yielding values between 7.5 and 12). In earlier versions of this paper, written prior to the circulation of Moreira and Moreira (2015), we instead followed Moreira and Moreira (2013) and set both tuning parameters equal to 1. Results based on this choice of tuning parameters may be found in the Supplemental Material.

relative performance of all the tests we consider at parameter values relevant for empirical practice, we calibrate our simulations based on data from Yogo (2004).[13] Yogo considered estimation of the elasticity of intertemporal substitution in eleven developed countries using linear IV and argued that estimation of this parameter appears to suffer from a weak instruments problem.[14] Yogo noted that both the strength of identification and the degree of heteroscedasticity appear to vary across countries, making his data set especially interesting for our purposes since it allows us to explore the behavior of the tests considered for a range of empirically relevant parameter values.

7.2.1. *Power Simulation Results*

We simulate the behavior of tests in the weak IV limit problem (6), and so require estimates for μ and Ω. To obtain these estimates, for each of the 11 countries in Yogo's data we calculate $\hat{\mu}$ and $\hat{\Omega}$ based on two-stage least squares estimates for the elasticity of intertemporal substitution, where $\hat{\Omega}$ is a Newey–West covariance matrix estimator using three lags.[15] A detailed description of this estimation procedure, together with the implementation of all tests considered, is given in the Supplemental Material. In particular, for the PI test we consider an estimator $\hat{\mu}_D$ which corresponds to the positive-part non-centrality estimator $\hat{r}_{\text{PP}}$ in the homoscedastic case discussed above.[16] The resulting power curves (based on 5,000 simulations for all tests) are plotted in Figures 1–2. Since for many countries the power curves are difficult to distinguish visually, in Table II we list the maximum regret for each test relative to the other tests studied, repeating the same exercise described above for the homoscedastic case.

Both the figures and the table highlight that while for many of the countries the K, QCLR, PI, MM, and QLR tests all perform well, as in the homoscedastic case there are some parameter values where the K test suffers from substantial declines in power relative to the other tests. In contrast to the homoscedastic case, the QCLR test does not fully resolve these issues. Instead, in cases where the K test exhibits especially large power declines, as in

[13]Note that simulation results on the power of the conditional QLR test in these calibrations are also reported in the supplement to I. Andrews and Mikusheva (2016a).

[14]The countries considered are Australia, Canada, France, Germany, Italy, Japan, the Netherlands, Sweden, Switzerland, the United Kingdom, and the United States. For comparability, we use Yogo's quarterly data for all countries, which in each case cover a period beginning in the 1970s and ending in the late 1990s, and take the endogenous regressor to be a risk-free interest rate.

[15]While the model assumptions imply that the GMM residuals $f_t(\beta)$ are serially uncorrelated at the true parameter value, the derivatives of the moment conditions $\frac{\partial}{\partial\beta} f_t(\beta)$ may be serially dependent.

[16]Specifically, we take $\hat{\mu}_D = D \cdot \sqrt{\max\{D'\Sigma_D^{-1}D - k, 0\}/D'\Sigma_D^{-1}D}$. When $\hat{\mu}_D = 0$, as suggested by footnote 9, we set $a_{\text{PI}}(D) = 1$.

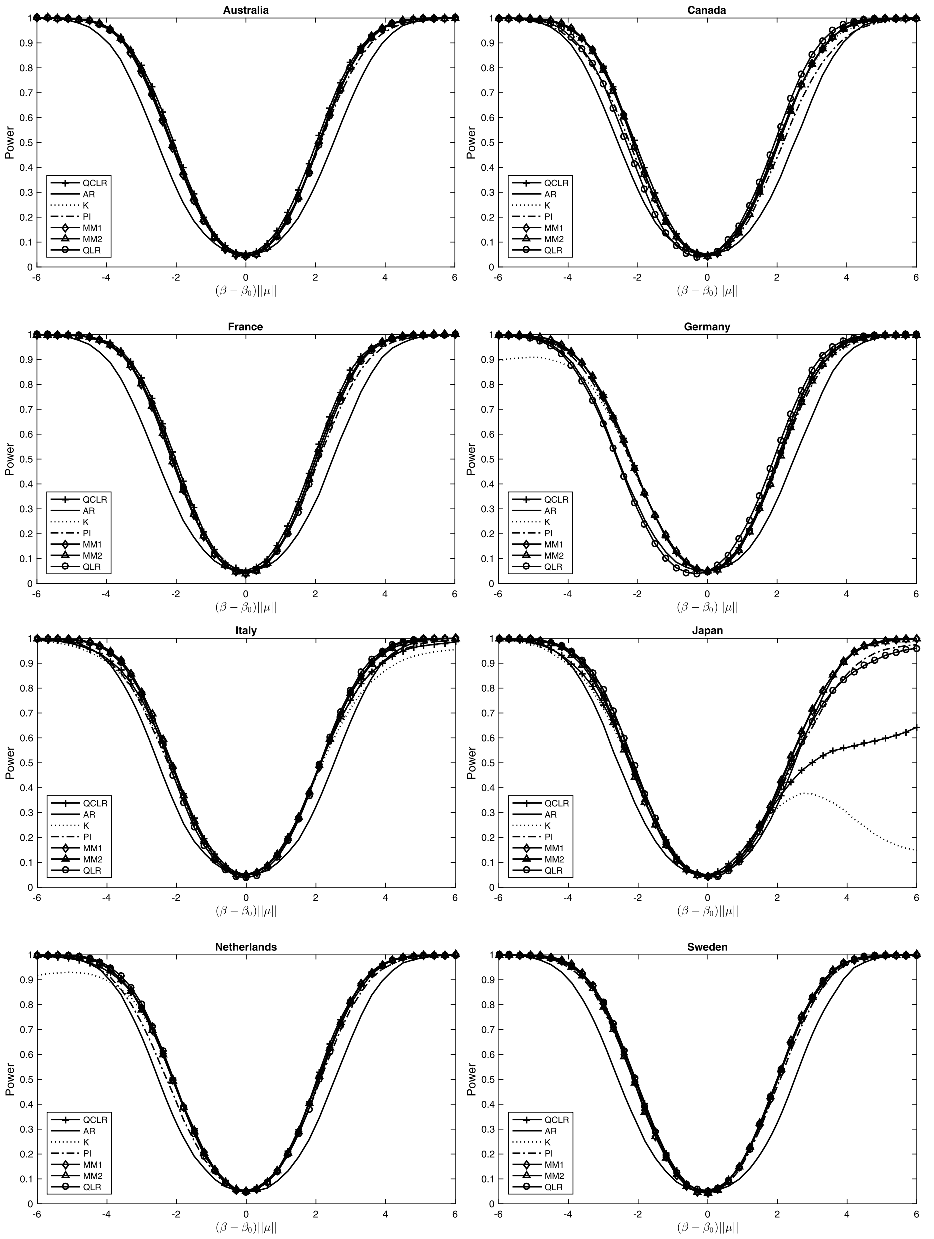

FIGURE 1.—Power functions for QCLR, AR (or S), K, PI, MM1-SU, MM2-SU, and QLR tests in simulation calibrated to Yogo (2004) data with four instruments.

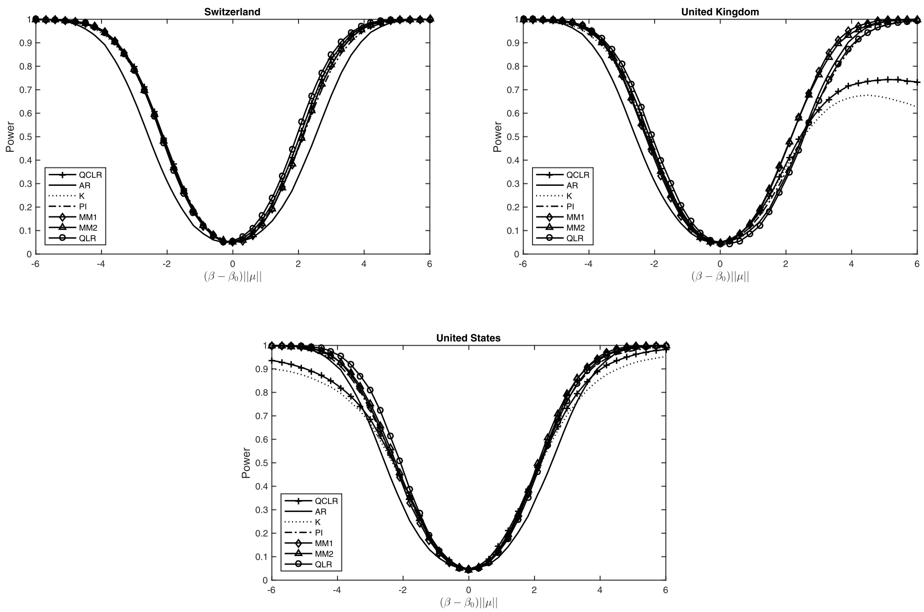

FIGURE 2.—Power functions for QCLR, AR (or S), K, PI, MM1-SU, MM2-SU, and QLR tests in simulation calibrated to Yogo (2004) data with four instruments.

TABLE II

MAXIMAL POINTWISE POWER SHORTFALL RELATIVE TO OTHER TESTS CONSIDERED, FOR SIMULATIONS CALIBRATED TO MATCH DATA IN YOGO (2004)[a]

	QCLR	AR	K	PI	MM1-SU	MM2-SU	QLR
Australia	4.06%	15.68%	6.28%	3.26%	1.68%	0.78%	1.72%
Canada	11.46%	20.98%	14.08%	10.06%	4.56%	4.88%	8.46%
France	5.16%	18.28%	5.48%	4.56%	2.12%	1.76%	2.60%
Germany	7.50%	20.90%	21.88%	6.94%	6.20%	6.38%	14.44%
Italy	9.86%	14.72%	14.14%	5.46%	1.50%	2.22%	3.38%
Japan	33.22%	16.56%	77.44%	7.32%	4.12%	4.66%	8.16%
Netherlands	10.34%	17.54%	16.88%	7.74%	1.62%	2.28%	2.28%
Sweden	4.56%	19.06%	4.78%	3.78%	1.92%	2.62%	1.04%
Switzerland	8.28%	21.36%	8.94%	7.86%	7.20%	6.96%	2.44%
United Kingdom	31.46%	18.86%	37.32%	12.88%	8.04%	7.04%	13.80%
United States	14.40%	17.22%	15.92%	8.74%	8.16%	6.36%	3.56%

[a]QCLR denotes the quasi-CLR test of Kleibergen (2005), while PI is the plug-in test discussed in Section 7.2.1. AR is the Anderson–Rubin (or S) test, K is Kleibergen's (2005) K test, and MM1-SU and MM2-SU are the weighted average power optimal SU tests of Moreira and Moreira (2013).

the simulations calibrated to match data from Japan and the United Kingdom, the QCLR test suffers from substantial power loss as well. While the QCLR test reduces power loss relative to the K test, the PI, MM1-SU, MM2-SU, and QLR tests do substantially better. While the power of the AR test is stable, for all countries its maximal power shortfall exceeds 10%.

The relatively poor performance of the QCLR test is driven by the fact, discussed above, that in the non-homoscedastic case the K statistic may focus on directions yielding low power. Since Kleibergen's QCLR test uses the CLR weight function, which is optimal in the homoscedastic case, it does not account for the fact that K may have worse performance when Σ lacks Kronecker product structure. In contrast, the PI test takes both the structure of Σ and the estimated value $\hat{\mu}_D$ into account when calculating $a_{\mathrm{PI}}(D)$, and so performs well in both the homoscedastic and non-homoscedastic cases. Additional (unreported) simulation results show that the PI test has power quite close to the infeasible MMRU test based on knowledge of the true μ_D, and the power of this MMRU test never exceeds that of the PI test by more than 2.6% in the designs considered here.

There is not a strict ranking among the PI, MM1-SU, MM2-SU, and QLR tests, and none of these tests has power dominating any of the others. Judged in terms of maximal power deficiency relative to the other tests considered, the MM2-SU test performs best, followed by the MM1-SU test, the PI test, and finally the QLR test.[17] As noted above, the PI test is unbiased, while the MM tests are locally unbiased and display no bias in these simulations. By contrast, as noted in I. Andrews and Mikusheva (2016a), the QLR test is not in general unbiased, and shows a small degree of bias in the calibration to German data. Overall, the four tests appear competitive, though the MM2-SU test has the smallest maximal power deficiency.[18]

8. CONCLUSION

This paper considers the problem of constructing powerful identification-robust tests for a broad class of weakly identified models. We show that tests which reject when a convex combination of the S and K statistics is large have a number of desirable power properties in a conditional problem. Restricting

[17]The choice of tuning parameters in the MM1-SU and MM2-SU tests is important for this result. If one instead uses tuning parameters as in Moreira and Moreira (2013), then the PI test has the smallest maximal power deficiency. Results under this alternative choice of tuning parameters may be found in the Supplemental Material.

[18]Additional simulation results based on calibrations to the Yogo data, using an equity return as the endogenous regressor rather than the risk-free rate, are reported in Moreira and Moreira (2015). Their simulation results did not consider the conditional QLR test but again found the PI test competitive with the MM tests, with power exceeding the MM1-SU test in several designs. They found the best overall performance for the MM2-SU test and argued that this in part reflects the benefit of allowing dependence on the data beyond the S, K, and D statistics.

attention to conditionally similar procedures which yield such convex combinations, we construct the class of conditional linear combination (CLC) tests. We show that CLC tests are unbiased, and further that the class of CLC tests is equivalent to an appropriately defined class of quasi-conditional likelihood ratio tests. To pick from the class of CLC tests, we suggest using MMR tests when feasible and PI tests when MMR tests are too difficult to compute. We show that PI tests match the near-optimal performance of the CLR test of Moreira (2003) in homoscedastic linear IV and are competitive with other recently proposed approaches in simulations calibrated to match an IV model with heteroscedastic time-series data.

REFERENCES

ANDERSON, T., 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. [2163]

ANDREWS, D. W. K., AND X. CHENG (2012): "Estimation and Inference With Weak, Semi-Strong, and Strong Identification," *Econometrica*, 80, 2153–2211. [2165]

——— (2013): "Maximum Likelihood Estimation and Uniform Inference with Sporadic Identification Failure," *Econometric Theory*, 173, 36–56. [2165]

ANDREWS, D. W. K., AND P. GUGGENBERGER (2014): "Asymptotic Size of Kleibergen's lm and Conditional lr Tests for Moment Condition Models," Cowles Foundation Working Paper. [2165]

ANDREWS, D. W. K., M. MOREIRA, AND J. STOCK (2006): "Optimal Two-Sided Invariant Similar Tests of Instrumental Variables Regression," *Econometrica*, 74, 715–752. [2157]

ANDREWS, I. (2016): "Supplement to 'Conditional Linear Combination Tests for Weakly Identified Models'," *Econometrica Supplemental Material*, 84, http://dx.doi.org/10.3982/ECTA12407. [2159]

ANDREWS, I., AND A. MIKUSHEVA (2016a): "Conditional Inference With a Functional Nuisance Parameter," *Econometrica*, 84, 1571–1612. [2157,2162,2165,2176,2177,2180]

——— (2016b): "A Geometric Approach to Weakly Identified Econometric Models," *Econometrica*, 84, 1249–1264. [2161,2165]

ELLIOTT, G., U. K. MÜLLER, AND M. WATSON (2015): "Nearly Optimal Tests When a Nuisance Parameter Is Present Under the Null Hypothesis," *Econometrica*, 83, 771–811. [2157,2158]

GUGGENBERGER, P., AND R. SMITH (2005): "Generalized Empirical Likelihood Estimators and Tests Under Partial, Weak, and Strong Identification," *Econometric Theory*, 21, 667–709. [2162-2164]

——— (2008): "Generalized Empirical Likelihood Ratio Tests in Time Series Models With Potential Identification Failure," *Journal of Econometrics*, 142, 134–161. [2163,2164]

GUGGENBERGER, P., J. RAMALHO, AND R. SMITH (2012): "Gel Statistics Under Weak Identification," *Journal of Econometrics*, 170, 331–349. [2162-2164]

KLEIBERGEN, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70, 1781–1803. [2163,2176]

——— (2005): "Testing Parameters in GMM Without Assuming They Are Identified," *Econometrica*, 73, 1103–1123. [2156,2157,2163,2164,2176,2179]

KOZIOL, J., AND M. PERLMAN (1978): "Combining Independent chi Squared Tests," *Journal of the American Statistical Association*, 73, 753–763. [2169]

KUBOKAWA, T., C. ROBERTS, AND A. SALEH (1993): "Estimation of Noncentrality Parameters," *The Canadian Journal of Statistics*, 21, 41–57. [2174]

MAGNUSSON, L. M. (2010): "Inference in Limited Dependent Variable Models Robust to Weak Identification," *Econometrics Journal*, 13, S56–S79. [2161]

MAGNUSSON, L., AND S. MAVROEIDIS (2010): "Identification-Robust Minimum Distance Estimation of the New Keynesian Phillips Curve," *Journal of Money, Banking, and Credit*, 42, 465–481. [2159,2161-2164]

MARDEN, J. (1982): "Combining Independent Noncentral chi Squared or f Tests," *The Annals of Statistics*, 10, 266–277. [2168]

MONTI, K., AND P. SEN (1976): "The Locally Optimal Combination of Independent Test Statistics," *Journal of the American Statistical Association*, 71, 903–911. [2169]

MONTIEL-OLEA, J. L. (2016): "Admissible Similar Tests: A Characterization," Unpublished Manuscript. [2157,2158]

MOREIRA, H., AND M. MOREIRA (2010): "Contributions to the Theory of Optimal Tests," Unpublished Manuscript. [2158]

——— (2013): "Contributions to the Theory of Optimal Tests," Unpublished Manuscript. [2176,2179,2180]

——— (2015): "Optimal Two-Sided Tests for Instrumental Variables Regression With Heteroskedastic and Autocorrelated Errors," Unpublished Manuscript. [2155,2157,2158,2176, 2180]

MOREIRA, M. (2001): "Tests With Correct Size When Instruments Can Be Arbitrarily Weak," Ph.D. Thesis, University of California, Berkeley. [2163]

——— (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71, 1027–1048. [2155,2156,2164,2171,2175,2176,2181]

MÜLLER, U. (2011): "Efficient Tests Under a Weak Convergence Assumption," *Econometrica*, 79, 395–435. [2157,2162,2175]

OTSU, T. (2006): "Generalized Empirical Likelihood Inference for Nonlinear and Time Series Models Under Weak Identification," *Econometric Theory*, 22 (22), 513–527. [2163]

RAMALHO, J., AND R. SMITH (2004): "Goodness of Fit Tests for Moment Condition Models," Unpublished Manuscript. [2163,2164]

SAXENA, K., AND K. ALAM (1982): "Estimation of the Non-Centrality Parameter of a chi Squared Distribution," *The Annals of Statistics*, 10, 1012–1016. [2174]

SMITH, R. (2007): "Weak Instruments and Empirical Likelihood: A Discussion of the Papers by D. W. K. Andrews and J. H. Stock and Y. Kitamura," in *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, Vol. 3. New York: Cambridge University Press, 238–260. Chapter 8. [2164]

STAIGER, D., AND J. STOCK (1997): "Instrumental Variables Regression With Weak Instruments," *Econometrica*, 65, 557–586. [2160]

STOCK, J., AND J. WRIGHT (2000): "Gmm With Weak Identification," *Econometrica*, 68, 1055–1096. [2156,2162,2163]

STOYE, J. (2009): "Minimax Regret," in *The New Palgrave Dictionary of Economics*, ed. by S. Durlauf and L. Blume. London, U.K.: Palgrave Macmillan. [2156]

YOGO, M. (2004): "Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak," *Review of Economics and Statistics*, 86, 797–810. [2157,2159,2177-2179]

Dept. of Economics, Massachusetts Institute of Technology, 50 Memorial Drive, E52-526, Cambridge, MA 02142, U.S.A.; iandrews@mit.edu.

Co-editor Elie Tamer handled this manuscript.

Manuscript received April, 2014; final revision received April, 2016.