Massachusetts Institute of Technology
Department of Economics
Working Paper Series

A Geometric Approach to Weakly Identified Econometric Models

Isaiah Andrews
Anna Mikusheva

Working Paper 12-15
May 29, 2012

Room E52-251
50 Memorial Drive
Cambridge, MA 02142

This paper can be downloaded without charge from the
Social Science Research Network Paper Collection at
http://ssrn.com/abstract=2070312

A Geometric Approach to Weakly Identified Econometric Models

Isaiah Andrews

Department of Economics

MIT

and

Anna Mikusheva

Department of Economics

MIT

May 29, 2012

A Geometric Approach to Weakly Identified Econometric Models

By Isaiah Andrews[1] and Anna Mikusheva[2]

Abstract

Many nonlinear Econometric models show evidence of weak identification, including many Dynamic Stochastic General Equilibrium models, New Keynesian Phillips curve models, and models with forward-looking expectations. In this paper we consider minimum distance statistics and show that in a broad class of models the problem of testing under weak identification is closely related to the problem of testing a "curved null" in a finite-sample Gaussian model. Using the curvature of the model, we develop new finite-sample bounds on the distribution of Anderson-Rubin-type statistics, which we show can be used to detect weak identification and to construct tests robust to weak identification. We apply the new method to a small-scale DSGE model and show that it provides a significant improvement over existing methods.

Key words: weak identification, statistical differential geometry

This draft: May 29, 2012.

1 Introduction

Empirical researchers in Economics frequently find that even in large samples the data provides little information about some model parameters. In such cases, known as weakly identified, the usual asymptotic approximations for estimators and test statistics may be quite poor, making standard approaches to inference unreliable. Weak identification has been detected in a wide range of *non-linear* estimation contexts, including estimation of the New Keynsian Phillips Curve (Dufour, Khalaf, and Kichian (2006), Kleibergen and Mavroeidis (2009b), Mavroeidis (2005), Nason and Smith (2008)), monetary policy rules (Mavroeidis (2010)), Dynamic Stochastic General Equilibrium (DSGE) Models (Ruge-Murcia (2007), Canova and Sala (2009), Iskrev (2010), I. Andrews and Mikusheva (2011), Guerron-Quintana, Inoue and Kilian (2009)), and Euler equations (Yogo (2004)).

[1]Department of Economics, M.I.T., 50 Memorial Drive, Building E52, Cambridge, MA, 02142. Email: iandrews@mit.edu. Financial support from the NSF Graduate Research Fellowship Program is gratefully acknowledged.

[2]Department of Economics, M.I.T., 50 Memorial Drive, Building E52, Cambridge, MA, 02142. Email: amikushe@mit.edu. Financial support from the Castle-Krob Career Development Chair is gratefully acknowledged. We are extremely grateful to Professor of Differential Geometry Toby Colding for discussing geometric issues with us, and for checking our geometric proofs. We are grateful to Gary Chamberlain, Victor Chernozhukov, Jerry Hausman, Mattias Cattaneo, Whitney Newey and Jim Stock for helpful comments.

The need for more reliable procedures robust to weak identification in non-linear contexts has inspired a large literature in econometrics - for a survey, see Dufour (2003) and Stock, Wright, and Yogo (2002).

A number of different testing procedures have been proposed in this literature, most of which address two situations: the case in which one is interested in testing the full parameter vector (also known as testing a simple hypothesis), and the case in which one is interested in testing only a subset of parameters but the parameters not under test (the nuisance parameters) are strongly identified. Examples of such tests include Stock and Wright (2000), Guggenberger and Smith (2005, 2008), Kleibergen (2005, 2007), I. Andrews and Mikusheva (2011), and Qu (2011). The literature to date has, however, been largely silent about the case in which part of the nuisance parameter vector may be weakly identified. A notable exception is the recent paper by D. Andrews and Cheng (2011).

Our paper directly addresses the question of inference with weakly identified nuisance parameters in the context of minimum distance estimation. We suggest a fully robust testing procedure which controls size without any assumption on the strength of identification of the parameters. Further, if the nuisance parameters are strongly identified, our procedure is asymptotically equivalent to the "concentrated out" S-test suggested by Stock and Wright (2000) for hypotheses with strongly identified nuisance parameters.

Our procedure is based on novel finite-sample bounds on the distribution of the test statistic under the null. For our asymptotics, we assume only that the model has reduced-form parameters which are asymptotically normal and hence we do not rely on any particular asymptotic embedding, such as those used by Stock and Wright (2000) or D. Andrews and Cheng (2011) to model weak identification. In fact, if one thinks a normal approximation to the distribution of the reduced-form estimates is reasonable, our bounds are finite-sample rather than asymptotic.

The bounds we derive rely on techniques from differential geometry which are new in the econometrics literature. Our starting point is the observation that hypotheses in non-linear models with strongly identified nuisance parameters are asymptotically linear in a geometrical sense. In contrast, hypotheses with weakly identified nuisance parameters need not be asymptotically linear and can exhibit substantial curvature even in large samples, leading to the breakdown of the usual asymptotic approximations. Our bounds can be viewed as a strengthening of the usual approximations, where rather

than appealing to asymptotic linearity of the null hypothesis we quantify the maximal deviation of the null from linearity and use it to construct stochastic bounds. As a result, under strong-identification asymptotics our bounds recover the usual approximations. The test we suggest uses a standard minimum-distance statistic paired with robust critical values derived using these geometric bounds. The validity of this approach relies only on assumptions which can be checked directly in applications, and our robust critical values are very easy to simulate. The bounds we derive are also of potential interest for a range of other applications, including testing nonlinear hypotheses and inference in highly non-linear models. Our approach differs from the statistical geometry literature initiated by Efron (1975) in that we produce finite sample bounds on the distribution of the test statistic, whereas the statistical geometry literature is primarily concerned with higher-order asymptotic approximations.

To date the dominant recommendation for testing hypotheses with weakly identified nuisance parameters has been the projection method (see Dufour and Jasiak (2001), Dufour and Taamouti (2005), Dufour, Khalaf, and Kichian (2006)). The strength of the projection method is that it requires no assumptions beyond the validity of the test for the full parameter vector. It is in general conservative, however, and may be extremely so in cases where the nuisance parameter is high-dimensional and/or strongly identified. Our approach is an improvement over the projection method, in that it uses the same test statistic paired with smaller critical values while still maintaining size.

If one knows that part of the nuisance parameter vector is strongly identified, it has been proved that in many cases one can obtain a more powerful test by concentrating out the nuisance parameter as in e.g. Stock and Wright (2000). Maintaining correct size in such cases, however, relies critically on the strong identification assumption on the nuisance parameter. In contrast, our approach requires no assumption of strong identification but, in the event that the nuisance parameters are strongly identified, is asymptotically equivalent to concentrating them out. In this sense, our robust critical values can be viewed as providing a continuous transition between projecting over and concentrating out the nuisance parameters, depending on the strength of identification.

We also use our bounds to derive a pre-test for weak identification which allows one to control size in two-step testing procedures. In particular, if one is deciding between concentrating out the nuisance parameter and using a robust procedure, we provide bounds on the curvature of the null hypothesis which can be used to guide the choice.

This procedure is quite different from existing tests for weak identification in nonlinear models such as Inoue and Rossi (2011), Iskrev (2010) and Wright (2003) in that the existing tests all effectively test the null of strong identification and hence do not control the probability of failing to detect weak identification. In contrast, our pre-test for weak identification directly bounds the finite-sample distribution of the test statistic.

As a side result, we provide an example showing that the distribution of an AR-type statistic in a nonlinear GMM model under weak identification is *not* in general dominated by the distribution of the same statistic under strong identification. Kleibergen and Mavroeidis (2009a: initial draft) claimed that such such stochastic dominance holds for some non-linear GMM models. If such a statement held universally (which Kleibergen and Mavroeidis did not claim), it would have meant that concentrating out the nuisance parameter controlled size regardless of identification strength. As we show in this paper, however, this dominance relationship does not in general hold for our setting, and concentrating out weakly identified nuisance parameters can yield substantial over-rejection.

We apply our approach to a small-scale DSGE model and find evidence of substantial curvature. We consider the problem of testing composite hypotheses about model parameters, and show that our robust critical values are substantially smaller than those used by the projection method while still controlling size.

The paper is structured as follows. In Section 2 we show that hypotheses with strongly identified nuisance parameters are asymptotically linear, while weakly identified nuisance parameters may cause non-trivial curvature of the null hypothesis. We also introduce several examples that fit our framework. In Section 3 we derive our geometric and stochastic bounds and introduce our fully robust test. In Section 4 we compare our testing procedures with existing methods, discuss the AR conjecture and introduce our pre-test. Section 5 is devoted to modified procedures for subsets of nuisance parameters, and Section 6 presents simulation results from applying our procedures to a small-scale DSGE model. All proofs may be found in the Appendix.

Throughout the paper we use the following notation: $\dot{\alpha}$ is the derivative of the function α, $\ddot{\alpha}$ is the second derivative, $B_R(x_0) = \{x \in \mathbb{R}^k : \|x - x_0\| \leq R\}$ is a k-dimensional ball of radius R with center x_0, and $B_R = B_R(0)$ is a ball around zero. Let $D_C = \{x = (x^{(1)}, x^{(2)}) : \|x^{(1)}\| \leq C, \|x^{(2)}\| \leq C, x^{(1)} \in \mathbb{R}^p, x^{(2)} \in \mathbb{R}^{k-p}\} \subset \mathbb{R}^k$, which is a natural generalization of a cylinder.

2 Model setting

Suppose we have a single observation $\hat{\theta} \sim N(\theta, I_k)$, where the true value of $\theta = \theta_0$ is unknown. Further, suppose we have a known regular p-dimensional manifold $\widetilde{S}$ in k-dimensional space. The assumption that $\widetilde{S}$ is a regular p-dimensional manifold essentially means that the tangent space to $\widetilde{S}$ at all points is a p-dimensional linear subspace.[3] The LR (likelihood ratio) statistic for testing a null hypothesis $H_0 : \theta \in \widetilde{S}$ is

$$LR = \min_{\theta \in \widetilde{S}} (\hat{\theta} - \theta)'(\hat{\theta} - \theta).$$

Example 1. Let $g : \mathbb{R}^k \to \mathbb{R}^{k-p}$ be a twice-continuously-differentiable function whose Jacobian has full rank at all points. Assume that we wish to test the null hypothesis $H_0 : g(\theta) = 0$. Then the set of points $\widetilde{S} = \{\theta \in \mathbb{R}^k : g(\theta) = 0\}$ describes a p-dimensional manifold which is known to the researcher. If g is a linear function, the null can be formulated as $H_0 : R\theta = r$, where R is $(k-p) \times k$ full-rank matrix and r is $(k-p) \times 1$ vector. In this special case, the manifold $\widetilde{S}$ described by the null hypothesis is a p-dimensional linear space.□

Example 2. Suppose we have a structural model which imposes that $\theta = \widetilde{m}(\beta)$ for some structural parameter β where the function $\widetilde{m} : \mathbb{R}^p \to \mathbb{R}^k$ is twice continuously differentiable with a full-rank Jacobian at all points. The image of the function $\widetilde{m}$ is a p-dimensional regular manifold $\widetilde{S}$ known to the researcher. Hence, the LR test for the hypothesis of correct specification will be based on the statistic

$$LR = \min_{\beta} (\hat{\theta} - \widetilde{m}(\beta))'(\hat{\theta} - \widetilde{m}(\beta)) = \min_{\theta \in \widetilde{S}} (\hat{\theta} - \theta)'(\hat{\theta} - \theta). \ \square$$

The distribution of the LR statistic under the null, that is for $\theta \in \widetilde{S}$, is in general non-standard and depends on the unknown nuisance parameter θ_0, the true value of θ, making inference difficult. In what follows, we derive bounds on distribution of LR for which it suffices to know the shape of the manifold $\widetilde{S}$; that is, which do not require knowledge of θ_0.

To proceed, it is useful for us to introduce the random vector $\xi = \hat{\theta} - \theta_0 \sim N(0, I_k)$ and the p-dimensional manifold $S = \{x : x = \theta - \theta_0, \theta \in \widetilde{S}\}$, which is simply $\widetilde{S}$ translated by $-\theta_0$. If the null is true, then the manifold S passes through the origin. The LR

[3] More details on regularity conditions can be found in section 3.

statistic is equal to the squared distance between ξ and S:

$$\rho(\xi, S)^2 = \min_{x \in S} (\xi - x)'(\xi - x). \tag{1}$$

The central statistical issue in this paper is how to characterize the distribution of $\rho(\xi, S)$, the squared distance from ξ to the manifold S, in terms which do not depend on the unknown θ_0. In particular, we develop bounds on this distribution that depend only on the curvature of the manifold $\widetilde{S}$, which is known to the researcher.

A well-known property of the normal distribution is that if S is a p-dimensional linear sub-space then the squared distance $\rho^2(\xi, S)$ has a χ^2_{k-p} distribution. Note that in this very special case the distribution does not depend on θ_0. Numerous classical results on testing in the presence of a nuisance parameter are based on this fact. Indeed, most of the classical statistics literature deals with testing hypotheses that are either linear or asymptotically linear, in the sense that S is either a linear subspace or arbitrarily well-approximated by one in large samples. In subsection 2.1 below, we argue that testing in the presence of strongly identified nuisance parameters is asymptotically equivalent to testing a linear hypothesis, while testing in the presence of weakly identified nuisance parameters tends to result in asymptotically non-linear null hypotheses. This has important implications for hypothesis testing, since if S is not a linear sub-space then the distribution of $\rho^2(\xi, S)$ is in general non-standard and depends on the whole shape of S.

There is one natural (less informative) bound that can be placed on $\rho^2(\xi, S)$ without any assumptions, namely that $\rho^2(\xi, S)$ is dominated by χ^2_k. Indeed, since $0 \in S$, we have

$$\rho(\xi, S)^2 = \min_{x \in S} (\xi - x)'(\xi - x) \leq (\xi - 0)'(\xi - 0) \sim \chi^2_k. \tag{2}$$

We argue below that this bound is precisely the one used by the "projection method", which is currently the main approach available for testing with weakly identified nuisance parameters. As discussed above, the major disadvantage of this bound is that it may yield quite conservative tests.

2.1 Weak identification is related to curvature

A parameter or group of parameters is known as weakly identified when it is point identified, but the data is not very informative about the true value. In such cases, it is

well known that many classical statistical approximations perform quite poorly. As a result, point estimates tend to be biased, many tests exhibit large size distortions, and confidence sets based on these tests have poor coverage. One framework used for developing better approximations in such cases is the drifting functions approach introduced by Stock and Wright (2000). This approach models weak identification using an asymptotic embedding in which the objective function is asymptotically flat along some directions. To fix ideas, let us consider a GMM model in which the moment function is separable in the data. In particular, assume that we observe a sample $\{x_i\}$ of size n consisting of identically and independently distributed observations such that

$$E(h(x_i) - M(\alpha, \beta)) = 0 \quad \text{for} \quad \alpha = \alpha_0, \ \beta = \beta_0. \tag{3}$$

Here $h(x)$ is a k-dimensional function with $E\|h(x_i)\|^4 < \infty$, while α and β are $k_\alpha \times 1$ and $k_\beta \times 1$ vectors respectively, for $k_\alpha + k_\beta \leq k$. Assume that $\theta_0 = (\alpha_0, \beta_0)$ is the unique point at which the moment condition (3) is satisfied, so that the model is point identified. As in Stock and Wright (2000), we can allow the function M to change as the sample size grows. In particular,

$$M(\alpha, \beta) = M_n(\alpha, \beta) = \widetilde{M}(\alpha) + \frac{1}{\sqrt{n}} M^*(\alpha, \beta), \tag{4}$$

where $\widetilde{M}(\alpha)$ and $M^*(\alpha, \beta)$ are fixed twice-continuously-differentiable functions with full-rank Jacobians. In this setting, α is strongly identified while β is weakly identified, because information about β does not accumulate as the sample size grows.

Suppose we are interested in testing a hypothesis about the structural parameters α and β. Consider first the problem of testing a full parameter vector hypothesis $H_0 : \alpha = \alpha_0, \beta = \beta_0$. To test this hypothesis, we can use a generalization of the AR (Anderson-Rubin) statistic introduced in Stock and Wright (2000):

$$AR(\alpha_0, \beta_0) = n \left(\frac{1}{n} \sum_i h(x_i) - M_n(\alpha_0, \beta_0) \right)' \Sigma^{-1} \left(\frac{1}{n} \sum_i h(x_i) - M_n(\alpha_0, \beta_0) \right),$$

where Σ is the covariance matrix of vector $h(x_i)$ (which we take to be nonsingular) or a consistent estimate thereof. Under the null, we have that $AR(\alpha_0, \beta_0) \Rightarrow \chi^2_k$. This result requires only that a central limit theorem hold for $h(x_i)$ and is fully robust towards

weak identification of β. An alternative test for the full parameter vector is suggested in Kleibergen (2005), which uses a score (LM) statistic.

To test a hypothesis with a strongly identified nuisance parameter, e.g. $H_0 : \beta = \beta_0$, we can use the AR statistic for a test on both α and β, minimized over the nuisance parameter α. In particular, we consider

$$AR(\beta_0) = \min_{\alpha} AR(\alpha, \beta_0).$$

Stock and Wright (2000) prove that under the null $AR(\beta_0) \Rightarrow \chi^2_{k-k_\alpha}$ regardless of the strength of identification of β. Interested readers may find a full proof of this result in Stock and Wright (2000): here, we instead show that the problem of testing $H_0 : \beta = \beta_0$ for a strongly identified nuisance parameter is asymptotically equivalent to that of testing a linear hypothesis in the Gaussian model described in the beginning of section 2.

To see that this is the case, define $\xi_n = \sqrt{n}\Sigma^{-1/2}(\frac{1}{n}\sum_i h(x_i) - M_n(\alpha_0, \beta_0))$. By the central limit theorem, $\xi_n \Rightarrow \xi \sim N(0, I_k)$. Let S_n be the image of the function

$$\begin{aligned} m_n(\alpha) =& \sqrt{n}\Sigma^{-1/2}(M_n(\alpha, \beta_0) - M_n(\alpha_0, \beta_0)) = \\ =& \sqrt{n}\Sigma^{-1/2}(\widetilde{M}(\alpha) - \widetilde{M}(\alpha_0)) + \Sigma^{-1/2}(M^*(\alpha, \beta_0) - M^*(\alpha_0, \beta_0)) = \\ =& \sqrt{n}\Sigma^{-1/2}(\widetilde{M}(\alpha) - \widetilde{M}(\alpha_0)) + O(||\alpha - \alpha_0||). \end{aligned} \tag{5}$$

The statistic $AR(\beta_0)$ is equal to $\rho^2(\xi_n, S_n)$. For any bounded set $\mathcal{B}$, the intersection $S_n \bigcap \mathcal{B}$ converges to the intersection of $\mathcal{B}$ with the k_α-dimensional linear sub-space S spanned by the columns of the Jacobian of $\widetilde{M}(\alpha)$ at point α_0. Indeed, according to equation (5) and the assumption that α is globally identified, one can easily show that the range of values of α such that $m_n(\alpha) \in S_n \bigcap \mathcal{B}$ is of order $1/\sqrt{n}$. Any regular manifold, however, is arbitrarily well approximated by its tangent space, which we can denote here by S, on an infinitesimal neighborhood of a regular point (see section 3.1 for definitions). As a result, it is easy to show that $\rho^2(\xi_n, S_n) \Rightarrow \rho^2(\xi, S) \sim \chi^2_{k-k_\alpha}$, where the last equality is true due to the fact discussed at the beginning of this section that the squared distance from a standard normal vector to a linear space passing through zero is χ^2-distributed. For another version of this asymptotic linearity result, see section 5, where we show that in models with strongly identified nuisance parameters the curvature is of order $1/\sqrt{n}$.

Tests for hypotheses with weakly identified nuisance parameters behave quite differently. In particular, we show that in general the curvature of a null hypothesis with a weakly identified nuisance parameter does not disappear asymptotically. To illustrate this point, assume that the hypothesis of interest is $H_0 : \alpha = \alpha_0$, so that β is a weakly identified nuisance parameter. Again, we consider the AR statistic minimized over the nuisance parameter:

$$AR(\alpha_0) = \min_{\beta} n \left(\frac{1}{n} \sum_i h(x_i) - M_n(\alpha_0, \beta) \right)' \Sigma^{-1} \left(\frac{1}{n} \sum_i h(x_i) - M_n(\alpha_0, \beta) \right).$$

Let us define $\xi_n = \sqrt{n}\Sigma^{-1/2}(\frac{1}{n}\sum_i h(x_i) - M_n(\alpha_0, \beta_0))$ as before, and let S_n be the image of the function

$$m_n(\beta) = \sqrt{n}\Sigma^{-1/2}(M_n(\alpha_0, \beta) - M_n(\alpha_0, \beta_0)) = \Sigma^{-1/2}(M^*(\alpha_0, \beta) - M^*(\alpha_0, \beta_0)).$$

By construction, S_n is a p-dimensional manifold in k-dimensional Euclidean space. In contrast to the strongly identified case, however, we have that S_n does not change with the sample size, so denote it S. Hence, if S_n is nonlinear for a given sample size, it remains nonlinear in the limit. As a result, we have that

$$AR(\alpha_0) = \rho^2(\xi_n, S) \Rightarrow \rho^2(\xi, S),$$

where $\xi \sim N(0, I_k)$ and S is a p-dimensional manifold, which is **not** in general a linear sub-space. Note, however, that this is precisely the problem discussed at the beginning of section 2. Hence, the problem of testing a hypothesis with weakly identified nuisance parameters using the AR statistic is asymptotically equivalent to that of testing a possibly nonlinear hypothesis in a Gaussian model. As a result, constructing bounds for the distribution of $\rho^2(\xi, S)$ will also allow us to conduct inference in models with weakly identified nuisance parameters.

Linearity vs strength of identification. We showed that the problem of testing a hypothesis with strongly identified nuisance parameters is asymptotically equivalent to that of testing that the mean belongs to linear subspace in a finite-dimensional Gaussian model. In contrast, if there are weakly identified nuisance parameters the manifold

corresponding to the null hypothesis need not converge to a linear subspace, so the usual critical values may be invalid. Rather than focusing on strength of identification, however, we may view the key distinction here as between linearity and non-linearity. In particular, while strong identification guarantees that the null hypothesis will correspond to a linear subspace in the limit, even with weakly identified nuisance parameters if S (the set of parameter values satisfying the null) happens to be a linear subspace, the usual $\chi^2_{k-k_\beta}$ limiting distribution will be correct. Hence, in models where the nuisance parameters enter the function M linearly, the usual (strong-identification) critical values for the AR statistic will yield asymptotically valid tests regardless of the strength of identification. Asymptotic linearity, not strong identification as such, is the essential condition. On a related note, Andrews and Mikusheva (2011) show that in a parametric model, a score test which concentrates out the nuisance parameter has asymptotically correct coverage even for testing a null hypothesis with a weakly identified nuisance parameter as long as this parameter enters the log-likelihood function linearly.

2.2 Minimal distance statistics

The analysis above generalizes to a broader minimum-distance context. Assume that we have a sample of size n from a model parameterized by structural parameters (α, β), which are related to reduced form parameters θ by $\theta = M_n(\alpha, \beta)$ where the function M_n is of the form described in equation (3). Further, suppose we have an estimator $\hat{\theta}$ of the reduced-form parameters which is consistent and asymptotically normal:

$$\sqrt{n}(\hat{\theta} - \theta) \Rightarrow N(0, \Sigma),$$

for Σ either known or consistently estimable: in the GMM example above, we could take $\theta = E[h(x_i)]$ and $\hat{\theta} = \frac{1}{n}\sum_{i=1}^n h(x_i)$. As before, α is strongly identified while β is weakly identified. The analysis now applies to so-called minimum distance statistics (which we will continue to refer to as AR for simplicity). In particular, to test the hypothesis $H_0 : \alpha = \alpha_0, \beta = \beta_0$ we use

$$AR(\alpha_0, \beta_0) = n\left(\hat{\theta} - M_n(\alpha_0, \beta_0)\right)' \Sigma^{-1} \left(\hat{\theta} - M_n(\alpha_0, \beta_0)\right),$$

while to test the hypothesis $H_0 : \beta = \beta_0$ with a strongly identified nuisance parameter we consider the statistic $AR(\beta_0) = \min_\alpha AR(\alpha, \beta_0)$ and pair it with $\chi^2_{k-k_\alpha}$ critical values. Taking $\xi_n = \sqrt{n}\Sigma^{-1/2}(\hat{\theta} - \theta_0)$, the same argument as above shows that this testing problem is asymptotically equivalent to testing a linear hypothesis in a Gaussian model. Likewise, testing a hypothesis with a weakly identified nuisance parameter is again asymptotically equivalent to testing a non-linear hypothesis in a Gaussian model.

Below we discuss several applied examples that can be cast into this setting.

2.2.1 Example: DSGE models

Dynamic Stochastic General Equilibrium (DSGE) models have recently been quite popular in applied Macroeconomics. These are highly non-linear, very multi-dimensional dynamic models describing the evolution of the main macro indicators in the economy and are used by many central banks. A number of concerns have been voiced about identification in these models (Ruge-Murcia (2007), Canova and Sala (2009), Iskrev (2011), I. Andrews and Mikusheva (2011), Guerron-Quintana, Inoue and Kilian (2009)), and many authors have noted that standard frequentist statistical procedures are unreliable. The source and extent of weak identification in such models is not well understood, and it is impossible to distinguish which parameters are weakly identified using currently-available procedures.

Several recent papers (Dufour, Khalaf and Kichian (2009), Guerron-Quintana, Inoue and Kilian (2009), I. Andrews and Mikusheva (2011), and Qu (2011)) suggest tests for full parameter vector hypotheses robust towards weak identification. With the exception of I. Andrews and Mikusheva (2011), these papers suggest the projection method for inference on subsets of parameters. Due to the high dimension of the parameter vector in many DSGE models, however, the projection method tends to be quite conservative.

Most DSGE models can be cast into our framework, which seems natural here, as one suggestion for how to estimate DSGE models is through two-step matching procedures (Christiano and Eichenbaum (1992), Rotemberg and Woodford (1997), Ruge-Murcia (2010)). Typical log-linearized DSGE models are of the form:

$$\Gamma_0(\beta) z_t = \Gamma_1(\beta) E_t z_{t+1} + \Gamma_2(\beta) z_{t-1} + \Gamma_3(\beta) u_t,$$

where z_t is a set of state variables at time t, β is a set of structural parameters, u_t are

i.i.d. mean zero shocks with identity covariance matrix, and $\Gamma_i(\beta)$ are known (often non-linear) functions. We observe variables $x_t = Cz_t$, where in many cases C is a deterministic matrix (usually a selection matrix). There are a number of known procedures for solving systems of linear rational expectation equations of this form- see Blanchard and Kahn (1980), Anderson and Moore (1985), King and Watson (1998), and Sims (2002) for different algorithms.

Once solved, the model can be written in the form

$$z_t = A(\beta)z_{t-1} + B(\beta)u_t,$$

where matrices $A(\beta)$ and $B(\beta)$ are generally nonlinear in β and solve the equations

$$(\Gamma_0 - \Gamma_1 A)A - \Gamma_2 = 0; \quad (\Gamma_0 - \Gamma_1 A)B - \Gamma_3 = 0.$$

In this context, a natural choice of reduced-form parameters is the auto-covariances of the observed vector-series x_t. In particular, let $\Sigma_x(j)$ be j-th order auto-covariance of x_t (for details see Iskrev (2010)):

$$\Sigma_x(j) = cov(x_t, x_{t-j}) = CA^j\Sigma_z(0)C'$$

where $\Sigma_z(0) = Ez_tz_t' = A\Sigma_z(0)A' + BB'$. It is helpful to write everything in vectorized form. In particular,

$$\theta_j(\beta) = vec(\Sigma_x(j)) = (C \otimes CA^j)(I - (A \otimes A))^{-1}vec(BB').$$

One may choose the reduced form parameter θ to be some subset of $vec(\Sigma_x(j))$, so $\theta = m(\beta) = W(\theta_0(\beta)', ..., \theta_j(\beta)')'$, where W is a selection matrix. In the absence of persistence (exact or near unit roots) the sample estimators

$$\hat{\theta}_j = vec\left(\frac{1}{T-j-1}\sum_{t=1}^{T-j}(x_{t+j} - \overline{x})(x_t - \overline{x})'\right)$$

of θ_j satisfy a central limit theorem and achieve normality quite quickly. As a result, normal approximations to the distribution of $\hat{\theta} = W(\hat{\theta}_0, ..., \hat{\theta}_j)$ are usually quite reliable for realistic sample sizes. Hence, we can conduct inference on the structural parameters

β using the AR statistic

$$AR(\beta) = n(\hat{\theta} - \theta(\beta))'\Sigma^{-1}(\hat{\theta} - \theta(\beta))$$

where Σ is the covariance matrix of the reduced-form parameter estimates.

2.2.2 Example: Phillips curve

Typical models of the New Keynesian Phillips curve relate present inflation to expectations about future inflation and past values of inflation. One such formulation is

$$\pi_t = \sum_{j=0}^{n} \lambda_j s_{t-j} + \gamma_f E_t \pi_{t+1} + \sum_{j=1}^{m} \gamma_j \pi_{t-j} + \varepsilon_t, \tag{6}$$

where π_t is the inflation at time t and s_t is some driving variable (e.g. labor costs). One way of estimating such models, popularized by Gali and Gertler (1999), is with instrumental variables regression.

Mavroeidis (2005) argues that models of forward-looking expectations (like the New Keynesian Phillips curve) require an analysis of identification distinct from the usual GMM-IV arguments considered in i.i.d. models. In particular, the error in the Phillips curve expression above will in general be autocorrelated and heterogeneous, raising identification issues above and beyond those faced in the i.i.d. case. Mavroeidis (2005) shows that the dynamics of the forcing variable s_t are extremely important for determining the identification of the model and that if these dynamics are insufficiently rich then the parameters in (6) will not be identified. He also argues that standard diagnostics for identification strength designed for i.i.d. models may be quite misleading when applied to Phillips curve estimation.

To cast New Keynesian Phillips curve estimation into our setting, suppose that the driving variable s_t is weakly exogenous and can be modeled as

$$s_t = \sum_{j=1}^{p} \rho_j s_{t-j} + \sum_{j=1}^{q} \phi_j \pi_{t-j} + v_t. \tag{7}$$

One can then solve the model described by (6) and (7) to obtain a reduced form solution

$$\pi_t = \sum_{j=0}^{l_s} \alpha_j s_{t-j} + \sum_{j=1}^{l_\pi} \delta_j \pi_{t-j} + \alpha_\varepsilon \varepsilon_t. \tag{8}$$

To test hypotheses on the structural parameters $\beta = (\lambda_0, ..., \lambda_n, \gamma_f, \gamma_1, ..., \gamma_m, \rho_1, ..., \rho_p, \phi_1, ..., \phi_q)$, we first estimate the reduced form parameters $\theta = (\alpha_0, ..., \alpha_{l_s}, \delta_1, ..., \delta_{l_\pi}, \rho_1, ..., \rho_p, \phi_1, ..., \phi_q)$ by OLS regressions (7) and (8). The function $\theta(\beta)$ connecting the structural parameters to the reduced form is given by the solution to the model and can be found in Mavroeidis (2005). Using this solution, we can test any hypothesis on the structural parameters by testing the implied hypothesis on the reduced form parameters. In particular, for a full parameter vector hypothesis $H_0 : \beta = \beta_0$ we consider the AR statistic:

$$AR(\beta_0) = n(\hat{\theta} - \theta(\beta_0))' \Sigma^{-1} (\hat{\theta} - \theta(\beta_0)),$$

where n is the sample size and Σ is a HAC-consistent estimator of the asymptotic variance of $\hat{\theta}$. Likewise, any other hypothesis about the structural parameters describes some manifold in the space of θ's and thus fits into the framework described in section 2.

3 Geometry

3.1 Manifolds, tangent spaces, curvature

In this paper we focus on regular manifolds embedded in k-dimensional Euclidean space with the usual Euclidean norm $\|\cdot\|$. A subset $S \subset \mathbb{R}^k$ is called a *p-dimensional regular manifold* if for each point $q \in S$ there exists a neighborhood V in $\mathbb{R}^k$ and a twice-continuously-differentiable map $\mathbf{x} : U \to V \bigcap S$ from an open set $U \subset \mathbb{R}^p$ onto $V \bigcap S \subset \mathbb{R}^k$ such that (i) $\mathbf{x}$ is a homeomorphism, which is to say it has a continuous inverse and (ii) the Jacobian $d\mathbf{x}_q$ has full rank. A mapping $\mathbf{x}$ which satisfies these conditions is called a parametrization or a system of local coordinates, while the set $V \bigcap S$ is called a coordinate neighborhood.

Note that the manifold S is defined as a set, rather than as a map. In keeping with this spirit, many of the statements below will be invariant to parametrization. Hence, if we have different parameterizations for the same manifold, which of them we use is

entirely a matter of convenience. In some problems it may be the case that there does not exist a global parametrization, that is a fixed mapping $\mathbf{x}$ satisfying the conditions above such that S is the image of $\mathbf{x}$. For instance, in Example 1 we defined S as the set of points q satisfying the $k-p$-dimensional restriction $g(q) = 0$ where the twice continuously differentiable function g had a Jacobian of rank $k-p$ at all points. Given this definition, the Implicit Function Theorem guarantees the existence of a local parametrization in a neighborhood of each point $q \in S$ but a global parametrization need not exist.

We begin by developing some geometrical concepts for the special case of a regular 1-dimensional manifold, also known as a curve. In particular, let S be a curve parameterized by $\alpha : (t_0, t_1) \to \mathbb{R}^k$ where α is twice continuously differentiable and (t_0, t_1) is an interval in $\mathbb{R}$. The *arc length* is defined as $s(t) = \int_{t_0}^{t} \|\dot{\alpha}(\tau)\| d\tau$. Without loss of generality, we can take α to be parameterized by arc length s, in which case at all points $\|\dot{\alpha}(s)\| = 1$ and the vector $\ddot{\alpha}(s)$ is perpendicular to $\dot{\alpha}(s)$. The vector $\dot{\alpha}(s)$ is called the tangent vector to S at $q = \alpha(s)$, while $\kappa(s) = \|\ddot{\alpha}(s)\|$ is called the *curvature* at q. The curvature measures how quickly the curve S deviates from its tangent line local to q, and the scaling is such that a circle of radius C has curvature $1/C$ at all points.

The change of variables from arbitrary parametrization t to the arc length s is not necessary for the calculation of curvature. In particular, as before let $\dot{\alpha}(t)$ and $\ddot{\alpha}(t)$ denote the first and second derivatives of α, now with respect to t. If we let $(\ddot{\alpha}(t))^{\perp}$ be the part of $\ddot{\alpha}(t)$ orthogonal to $\dot{\alpha}(t)$, then the curvature at $q = \alpha(t)$ is $\kappa(t) = \frac{\|(\ddot{\alpha}(t))^{\perp}\|}{\|\dot{\alpha}(t)\|^2}$. One can show that this definition of curvature is invariant to parametrization, and hence that in the special case of a curve parameterized by arc length it reduces to the definition given above.

These concepts can all be extended to general regular manifolds. Fixing a p-dimensional manifold S, for any curve $\alpha : (-\varepsilon, \varepsilon) \to S$ on S which passes through the point $q = \alpha(0) \in S$ the tangent vector $\dot{\alpha}(0)$ is called a tangent vector to S at q. For $\mathbf{x}$ a system of local coordinates at q, the set of all tangent vectors to S at q coincides with the linear space spanned by the gradient $d\mathbf{x}_q$ and is called the *tangent space* to S at q (denoted $T_q(S)$). While we have defined the *tangent space* using the local coordinates $\mathbf{x}$, as one would expect from its geometrical interpretation $T_q(S)$ is independent of the parametrization.

To calculate the curvature at q, consider a curve $\alpha : (t_0, t_1) \to S$ which lies in S and passes through $q = \alpha(0)$. Taking $T_q^{\perp}$ to be the $k-p$-dimensional linear space orthogonal

to $T_q(S)$ and $X = \dot{\alpha}(0) \in T_q(S)$ to be the tangent vector to α at q, define

$$\kappa_q(X, S) = \frac{\left\|(\ddot{\alpha}(0))^{\perp}\right\|}{\|X\|^2},$$

where $(W)^{\perp}$ stands for the projection of W onto the space $T_q^{\perp}$. One can show that $\kappa_q(X, S)$ depends on the curve α only through X. The measure of curvature we consider is

$$\kappa_q(S) = \sup_{X \in T_q(S)} \kappa_q(X, S) = \sup_{X \in T_q(S)} \frac{\left\|(\ddot{\alpha}(0))^{\perp}\right\|}{\|X\|^2}. \tag{9}$$

This measure of curvature is closely related to the Second Fundamental Tensor (we refer the interested reader to Kobayashi and Nomizu (1969, v.2, ch. 7)), and is equal to the maximal curvature over all geodesics passing through the point q. As with the curvature measure discussed for curves, (9) is invariant to the parametrization. Also analogous to the 1-dimensional case, if S is a p-dimensional sphere of radius C then for each $q \in S$ we have $\kappa_q(S) = 1/C$. Finally, if S is a linear subspace then its curvature is zero at all points.

How to calculate curvature in practice. Let S be a p-dimensional manifold in $\mathbb{R}^k$, and let $\mathbf{x}$ be a local parametrization at a point q, $q = \mathbf{x}(y^*)$. Denote the derivatives of $\mathbf{x}$ at q by $v_i = \frac{\partial \mathbf{x}}{\partial y_i}(y^*)$. By the definition of a local parametrization, we know that the Jacobian $Z = (v_1, ..., v_p)$ is full rank, so the tangent space $T_q(S) = span\{v_1, ..., v_p\}$ is p-dimensional. As before, for any vector $W \in \mathbb{R}^k$ let $W^{\perp}$ denote the part of W orthogonal to $T_q(S)$, that is, $W^{\perp} = (I - Z(Z'Z)^{-1}Z')W$. Finally, denote the p^2 vectors of second derivatives $V_{ij} = \frac{\partial^2}{\partial y_i \partial y_j}\mathbf{x}(y^*)$. The curvature can then be written as

$$\kappa_q(S) = \sup_{\substack{u=(u_1,..,u_p)\in\mathbb{R}^p \\ \|\sum_{i=1}^p u_i v_i\|=1}} \left\| \sum_{i,j=1}^{p} u_i u_j V_{ij}^{\perp} \right\| = \sup_{(w_1,...,w_p)\in\mathbb{R}^p} \frac{\left\|\sum_{i,j=1}^p w_i w_j V_{ij}^{\perp}\right\|}{\left\|\sum_{i=1}^p w_i v_i\right\|^2}. \tag{10}$$

3.2 Geometric bounds

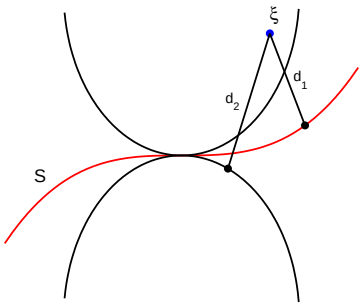
In this section we establish a bound on the distribution of the distance in $\mathbb{R}^k$ from a random point $\xi \sim N(0, I_k)$ to a p-dimensional non-random manifold S that contains

zero. Our bound depends on the maximal curvature $\kappa_q(S)$ over all relevant points in the manifold S. Our bound will be based on global properties of the manifold, in the sense of properties that hold on a fixed bounded set, but we abstract from the behavior of the manifold at infinity as irrelevant. In what follows, we restrict attention to a connected part of the manifold that lies inside of a (large) finite cylinder centered at zero.

We derive our bound in two steps: first, we construct an envelope for the manifold S using a collection of p-dimensional spheres. We show that the distance from any point ξ to S is bounded above by the distance from ξ to the most distant sphere in the collection we consider. Second, we show that our geometric construction implies a bound on the distribution of $\rho^2(\xi, S)$ and hence on the distribution of AR (or LR) statistics. To provide intuition for our main statement we walk the reader through two simple cases in which the construction of the envelope can be easily visualized.

Case 1 (k=2, p=1): A curve in $\mathbb{R}^2$. Consider a curve S passing through zero (i.e. $(0,0) \in S$). Suppose that the curvature of S is less than or equal to $1/C$ for all points in S. If we imagine two circles of radius C tangent to S at zero, we can see that the curve lies between them- see Figure 1 for illustration. Since S lies between the circles, the distance from any point ξ to S (denoted by d_1 in Figure 1) does not exceed the distance from ξ to the further of the two circles (denoted by d_2). This is the geometrical bound we use. Note that if the maximal curvature of S goes to zero at all points (so that $C \to \infty$)

Figure 1: Bounding a line between two circles.



then the two bounding circles converge to the tangent line to S at zero on any bounded set. Further, note that the distribution of the distance d_2 from a normal random vector to the furthest of two circles depends only on C and is easy to simulate.

The logic of this example is quite straightforward to generalize to the case of a $k-1$-

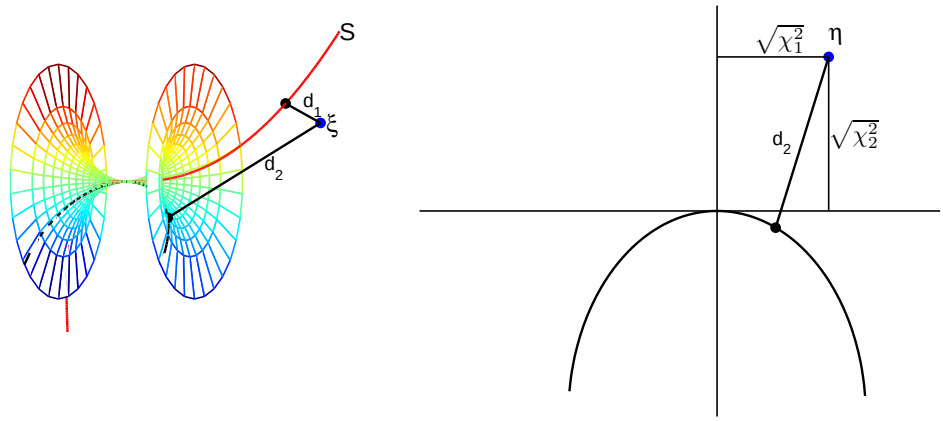


Figure 2: Left panel: the envelope for a space curve in $\mathbb{R}^3$. Right panel: distribution of the distance d_2

dimensional manifold in $\mathbb{R}^k$, known as a hyper-surface or a manifold of co-dimension 1.[4] If a regular $k-1$ dimensional manifold S in $\mathbb{R}^k$ has curvature $\kappa_q(S) \leq 1/C$ at all $q \in S$, consider the two $k-1$-dimensional spheres of radius C which are tangent to the manifold at zero. One can show (see Theorem 1 below) that as in the one-dimensional case S lies between these two spheres. Hence, we again have that the distance from any point ξ to S is bounded above by the distance from ξ to the furthest of the two spheres. Likewise, if the maximal curvature of S goes to zero (so that $C \to \infty$) we again have that on any bounded set the two spheres converge to the tangent space to S at zero, which in this case is a $k-1$-dimensional hyperplane.

Dealing with manifolds of co-dimension greater than 1 is much more challenging, but the basic principle of the approach can be illustrated using a curve in $\mathbb{R}^3$.

Case 2 (k=3, p=1): A curve in $\mathbb{R}^3$. Suppose now that we have a one-dimensional space curve S in $\mathbb{R}^3$ which passes through zero and whose curvature at all points is bounded above by $1/C$. We construct our envelope by considering the collection of all one-dimensional circles of radius C tangent to S at zero. Equivalently, one can take a given circle tangent to S at zero and rotate it around the tangent line. An example of the resulting surface is given on the left panel of Figure 2: as in the case of co-dimension 1, we can see that the curve S lies inside the envelope. One can show that the distance from any point ξ to the curve S (denoted by d_1 in Figure 2) is bounded above by the distance from ξ to the furthest circle in the collection used to construct the envelope

[4]The co-dimension of a manifold is the difference between the dimension of the space and the dimension of the manifold.

(denoted d_2). Note that if the curvature of S goes to zero at all points (so that $C \to \infty$) then on any bounded set the envelope we consider converges to the tangent line to S at zero.

This geometric bound immediately implies a bound on the distribution of $\rho(\xi, S)$. For $\xi \sim N(0, I_3)$ the distribution of the distance d_2 from ξ to the furthest circle is quite simple to simulate. One can show that it is distributed as the distance from two-dimensional random vector η depicted on the right panel of Figure 2 to the circle of radius C with center $(0, -C)$ where the coordinates of η are distributed as independent $\sqrt{\chi_1^2}$ and $\sqrt{\chi_2^2}$ random variables.

General case With the intuition provided by these examples, we now turn to the general case. Let S be a regular connected p-dimensional manifold in $\mathbb{R}^k$ passing through zero. By the rotation invariance of standard normal vectors we can assume without loss of generality that the tangent space $T_0(S)$ to manifold S at zero is spanned by first p basis vectors. For each $x \in \mathbb{R}^k$, let $x = (x^{(1)}, x^{(2)})$ where $x^{(1)} = (x_1, ..., x_p) \in \mathbb{R}^p$ contains the first p coordinates of x while $x^{(2)} = (x_{p+1}, ..., x_k) \in \mathbb{R}^{k-p}$ contains the last $k - p$. In what follows, we restrict attention to points on the manifold that lie inside of a (large) finite cylinder $D_C = \{x = (x^{(1)}, x^{(2)}) : \|x^{(1)}\| \leq C, \|x^{(2)}\| \leq C, x^{(1)} \in \mathbb{R}^p, x^{(2)} \in \mathbb{R}^{k-p}\} \subset \mathbb{R}^k$. Let S_C be the intersection $S \bigcap D_C$ if it is connected or the *connected* part of $S \bigcap D_C$ that passes through zero (that is, the part of $S \bigcap D_C$ which can be reached by continuous paths lying in $S \bigcap D_C$ which pass through zero) if $S \bigcap D_C$ is not connected. Note that $\rho(\xi, S) \leq \rho(\xi, S_C)$.

To obtain some of our bounding results, we need one further assumption:

Assumption 1 *For any $y^{(1)} \in \mathbb{R}^p$ with $\left\|y^{(1)}\right\| \leq C$ there exists a point $x \in S_C$ such that $x^{(1)} = y^{(1)}$.*

Assumption 1 requires that the projection of S on the tangent space to S at zero covers some fixed p-dimensional ball, and hence that S has dimension p in a global sense. By a local property we mean one that holds on an infinitesimal neighborhood of a point. In contrast, by a global property we mean one that holds on a fixed bounded set. We have already imposed a local dimensionality assumption on S by restricting the rank of the tangent space at all points. The distribution of the AR statistic, however, depends on global properties of the manifold S and so to bound the distribution we need a

global dimensionality assumption. To illustrate why local dimensionality assumptions are insufficient, imagine a strip $S = \{(x, y, z) \in \mathbb{R}^3 : z = 0, -\varepsilon < y < \varepsilon, x \in \mathbb{R}\}$ in $\mathbb{R}^3$. At any point $q \in S$ the dimension of the tangent space is equal to 2, but if $\varepsilon > 0$ is small enough then S does not satisfy Assumption 1. For ε sufficiently small, however, the distance from ξ to S behaves like the distance from ξ to the line $S^* = \{(x, y, z) : y = 0, z = 0\}$, which is one dimensional both locally and globally.

Theorem 1 *Let S be a regular p-dimensional manifold in $\mathbb{R}^k$ passing through zero. Assume that the tangent space $T_0(S)$ is spanned by first p basis vectors. Assume that for some constant $C > 0$ we have that $\kappa_q(S) < \frac{1}{C}$ for all points $q \in S_C$. Then:*

(a) Manifold S_C lies inside the set $\mathcal{M} \cap D_C$, where

$$\mathcal{M} = \{\|x^{(1)}\|^2 + (C - \|x^{(2)}\|)^2 \geq C^2\}. \tag{11}$$

(b) If Assumption 1 is satisfied, then for any point $\xi \in \mathbb{R}^k$ we have

$$\rho(\xi, S) \leq \max_{u \in \mathbb{R}^{p-k}, \|u\|=1} \rho(\xi, N_u),$$

where $N_u = \{x \in \mathbb{R}^k : x = (x^{(1)}, zu), x^{(1)} \in \mathbb{R}^p, z \in \mathbb{R}_+, \|x^{(1)}\|^2 + (C - z)^2 = C^2\}$.

(c) $\max_{u \in \mathbb{R}^{p-k}, \|u\|=1} \rho(\xi, N_u) = \rho(\xi, N_{\widetilde{u}})$, where $\widetilde{u} = -\frac{1}{\|\xi^{(2)}\|}\xi^{(2)}$.

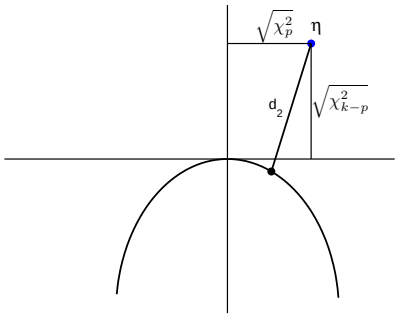
(d) If $\xi \sim N(0, I_k)$ we have for all x, y:

$$P\left\{\max_{u \in \mathbb{R}^{p-k}, \|u\|=1} \rho^2(\xi, N_u) \leq x, \|\xi\| \leq y\right\} = P\left\{\rho_2^2(\eta, N_2^C) \leq x, \|\eta\| \leq y\right\},$$

where the coordinates of the 2-dimensional random vector $\eta = (\sqrt{\chi_p^2}, \sqrt{\chi_{k-p}^2}) \in \mathbb{R}^2$ are independently distributed, $N_2^C = \{(z_1, z_2) \in \mathbb{R}^2 : z_1^2 + (C + z_2)^2 = C^2\}$ is a circle of radius C with the center at $(0, -C)$, and ρ_2 is Euclidian distance in $\mathbb{R}^2$.

Theorem 1 (a) establishes that the manifold S_C lies inside the set $\mathcal{M}$ bounded by an envelope we construct from a collection of p-dimensional spheres N_u. Statement (b) asserts that the distance from a point ξ to the manifold S is bounded by the distance from ξ to the furthest sphere in this collection, while (c) picks out exactly which sphere $N_{\widetilde{u}(\xi)}$ is the furthest away for a given ξ. Finally, (d) shows that the distribution of the

Figure 3: The stochastic bound described in Theorem 1 (d).



distance from $\xi \sim N(0, I_k)$ to $N_{\tilde{u}(\xi)}$ is the same as the distribution of the distance from a random variable η to a particular circle in $\mathbb{R}^2$ as depicted in Figure 3.

3.3 Stochastic bound

Theorem 1 implies a bound on the distribution of the distance from $\xi \sim N(0, I_k)$ to a p-dimensional manifold S. Assume that for some $C > 0$, S satisfies all the assumptions of Theorem 1 including Assumption 1. Then almost surely,

$$\rho^2(\xi, S) \leq \rho^2(\xi, N_{\tilde{u}}), \tag{12}$$

as follows from statement (b) of Theorem 1. By Theorem 1 (d), the distribution of the right hand side of (12) is the same as the distribution of the random variable variable ψ_C defined as

$$\psi_C = \rho_2^2(\eta, N_2^C), \tag{13}$$

where the coordinates of the two-dimensional random vector $\eta = (\sqrt{\chi^2_p}, \sqrt{\chi^2_{k-p}}) \in \mathbb{R}^2$ are independently distributed, $N_2^C = \{(z_1, z_2) \in \mathbb{R}^2 : z_1^2 + (C + z_2)^2 = C^2\}$ is a circle of radius C with the center at $(0, -C)$, and ρ_2 is Euclidean distance in $\mathbb{R}^2$. Combining these results, we establish the bound

$$P\left\{\rho^2(\xi, S) \geq x\right\} \leq P\left\{\psi_C \geq x\right\} \qquad \text{for all } x > 0,$$

so the distribution of ψ_C is an upper bound on the distribution of $\rho^2(\xi, S)$. We make the following observations:

(1) The distribution of ψ_C depends only on the dimension of the space k, the dimension of the manifold p and the maximal value of the curvature, $\frac{1}{C}$.

(2) The distribution of ψ_C is stochastically increasing in the maximal curvature and hence stochastically decreasing in C, so if $C_1 < C_2$ then ψ_{C_1} first-order stochastically dominates ψ_{C_2}.

(3) $\psi_C \Rightarrow \chi^2_{k-p}$ as $C \to \infty$, so if the curvature converges to zero at all relevant points then our bounding distribution converges to distribution of the distance from $\xi \sim N(0, I_k)$ to a p-dimensional linear subspace.

(4) At the other extreme, $\psi_C \Rightarrow \chi^2_k$ as $C \to 0$ so if the curvature of the manifold becomes arbitrarily large our bound coincides with the naive bound (2) that can be imposed without any assumptions on the manifold.

We want to emphasize that what we suggest is a stochastic bound that holds under quite general assumptions. If the model of interest has additional structure, this can potentially be exploited to obtain tighter bounds.

3.4 Statistical application of the stochastic bound

Suppose we have a single observation $\widehat{\theta}$ from a population $\widehat{\theta} \sim N(\theta_0, \Sigma)$ with an unknown mean θ_0. We wish to test a hypothesis of the form $H_0 : \theta_0 = \theta(\beta)$ for some value of the p-dimensional structural parameter $\beta \in U \subset \mathbb{R}^p$. As discussed above, this problem may arise when we have asymptotically normal reduced form estimates and are interested in testing hypotheses on the structural parameters as in sections 2.2.1 and 2.2.2. We use the AR statistic

$$AR = \min_{\beta}(\widehat{\theta} - \theta(\beta))'\Sigma^{-1}(\widehat{\theta} - \theta(\beta)) = \rho^2(\xi, S), \tag{14}$$

where $\xi = \Sigma^{-1/2}(\widehat{\theta} - \theta_0) \sim N(0, I_k)$ is an appropriately normalized version of our reduced form estimate $\hat{\theta}$ and the manifold $S = \{\Sigma^{-1/2}(\theta(\beta) - \theta_0), \beta \in U \subset \mathbb{R}^p\} \subset \mathbb{R}^k$ describes the restrictions imposed on the reduced-form parameters by the tested hypothesis and passes through zero if the null is true. If the manifold S satisfies the assumptions of Theorem 1 then by the argument in Section 3.3 the AR statistic is stochastically dominated by ψ_C

under the null, so if we use the $(1-\alpha)-$quantile of the distribution of ψ_C (which is easy to simulate) as a critical value the resulting test has size not exceeding α.

A practical question is what value of C to use. According to Theorem 1, the value of C is tied to the maximum of the curvature of S over the intersection of S with a cylinder D_C centered at zero. Notice, however, that in practice we do not observe the manifold S, since it depends on the unknown θ_0. However, the desired curvature is the same as the maximal curvature of the manifold $\widetilde{S} = \{\Sigma^{-1/2}\theta(\beta), \beta \in U \subset \mathbb{R}^p\} \subset \mathbb{R}^k$ over all points in the intersection of $\widetilde{S}$ with the cylinder $\widetilde{D}_C(x_0) = \{x \in \mathbb{R}^k : x - x_0 \in D_C\}$ centered at $x_0 = \Sigma^{-1/2}\theta_0$. This maximal curvature, in turn, is clearly bounded above by the maximal curvature over the whole manifold, so if we take $\widetilde{C} = 1/\left(\max_{q\in\widetilde{S}} \kappa_q(\widetilde{S})\right)$, using critical values based on $\psi_{\widetilde{C}}$ provides a test that controls size. Moreover, since $\widetilde{C}$ does not depend on any unobservables, a test based on these critical values is feasible.

If the null hypothesis has a global parametrization, as when $H_0 : \theta_0 = \theta(\beta), \beta \in U$, let $\kappa(\beta) = \kappa_{q=\theta(\beta)}(S)$. The latter is a function on U which depends only on the first two derivatives of $\theta(\beta)$. Hence, if we can evaluate these derivatives finding $\widetilde{C} = \max_\beta \kappa(\beta)$ is a standard non-stochastic optimization problem. If $\theta(\beta)$ is fairly tractable we may be able to solve for $\widetilde{C}$ analytically, while if not we can use the usual menu of numerical optimization techniques, such as Newton's method.

There are a variety of problems, however, in which using $\widetilde{C}$ may be unappealing. For example, it may be that calculating derivatives of $\theta(\beta)$ is challenging, or that the manifold has irregularities or points of high curvature which are far away from $\widehat{\theta}$. In such cases we may wish to restrict attention to the curvature of the manifold over some smaller set, which raises two issues. First, we do not know the true value θ_0 and hence the center of the cylinder $\widetilde{D}_C(x_0)$. Second, if the manifold is close to flat (so C is large) to find the maximal curvature over $\widetilde{D}_C(x_0)$ we might need to check the curvature over a huge set, which could be very computationally demanding.

We suggest a test which overcomes both of these problems and is easy to implement in practice. For a fixed value R, let $C \wedge R = \min\{C, R\}$. Denote by $F_\alpha(C, R, k, p)$ the $\alpha-$quantile of the distribution of $\psi_C(R)$ defined as

$$\psi_C(R) = \begin{cases} \rho_2^2(\eta, N_2^C) & \text{if } \|\eta\| \leq R; \\ \|\eta\|^2 & \text{if } \|\eta\| > R, \end{cases}$$

where η and and N_2^C are defined in statement (d) of Theorem 1. For any finite R the distribution of $\psi_C(R)$ provides a weaker bound than the distribution of ψ_C. This is the price paid for calculating curvature over a smaller set of points.

Lemma 1 *Assume that we have a single observation $\widehat{\theta}$ from a population $\widehat{\theta} \sim N(\theta_0, \Sigma)$ with unknown mean θ_0. We wish to test the hypothesis $H_0 : \theta_0 = \theta(\beta)$ for some $\beta \in U \subset \mathbb{R}^p$. Let $\widetilde{S} = \{\Sigma^{-1/2}\theta(\beta), \beta \in U \subset \mathbb{R}^p\} \subset \mathbb{R}^k$ be a regular p-dimensional manifold, and $\widetilde{B} = B_{(1+\sqrt{2})R}(\widehat{x})$ a ball of radius $(1+\sqrt{2})R$ around $\widehat{x} = \Sigma^{-1/2}\widehat{\theta}$, where R is such that $P\{\chi_k^2 \geq R^2\} < \alpha$. Let*

$$\widehat{C} = \begin{cases} \left(\min_{q \in \widetilde{S} \bigcap \widetilde{B}} 1/\kappa_q(\widetilde{S})\right) \wedge R, & \text{if } \widetilde{S} \bigcap \widetilde{B} \neq \emptyset; \\ 0, & \text{if } \widetilde{S} \bigcap \widetilde{B} = \emptyset. \end{cases}$$

Assume that for any $x \in \widetilde{S}$ such that $\|x - \widehat{x}\| \leq R$ we have that the projection of $\widetilde{S} \bigcap B_R(x)$ onto $T_x(\widetilde{S})$ contains a p-dimensional ball centered at x with radius $\widehat{C} \wedge R$. Then the test which rejects the null if and only if $AR = \min_\beta (\widehat{\theta} - \theta(\beta))' \Sigma^{-1} (\widehat{\theta} - \theta(\beta)) > F_{1-\alpha}(\widehat{C}, R, k, p)$ has size not larger than α.

4 Comparison with other methods available for testing hypotheses with weak nuisance parameters

As previously discussed, there is a wide literature devoted to the problem of weak-identification-robust tests for the full parameter vector and for hypotheses with strongly identified nuisance parameters, but much less is known about testing with weakly identified nuisance parameters.

Projection method. Recently, the projection method has been the standard approach to inference with weakly identified nuisance parameters. The projection method was introduced and popularized in econometrics by Dufour and Jasiak (2001) and Dufour and Taamouti (2005), and recent applications to non-standard testing problems in econometrics include Dufour, Khalaf, and Kichian (2006), Guerron-Quintana, Inoue and Kilian (2009), and Qu (2011).

The projection method is based on the observation that the hypothesis $H_0 : \alpha = \alpha_0$

with nuisance parameter β is equivalent to the hypothesis

$$H_0 : \exists \ \ \beta_0 \text{ s.t. } \alpha = \alpha_0, \beta = \beta_0.$$

Hence, to test a hypothesis on α we can use the statistic $AR(\alpha_0) = \inf_{\beta^*} AR(\alpha_0, \beta^*)$, and since

$$AR(\alpha_0) = \inf_{\beta^*} AR(\alpha_0, \beta^*) \leq AR(\alpha_0, \beta_0) \sim \chi^2_k,$$

we know that using χ^2_k critical values will yield a test which controls size. The name "projection method" stems from the fact that constructing confidence sets for α with this procedure is equivalent to constructing a joint confidence set for (α_0, β_0) using the full-vector AR statistic and then projecting this set on the parameter space for α.

The obvious advantage of the projection method is that it requires no assumptions about the strength of identification of β, since it relies only on the validity of the test for the full parameter vector. Other advantages include that it is quite easy to use and very broadly applicable. The primary disadvantage of the projection method is its conservativeness. Our test, introduced in Section 3.4, is based on the same statistic as the projection method (AR minimized over the nuisance parameters) but uses smaller critical values while still maintaining size. Only in the limiting case of infinitely high curvature ($C = 0$) do our critical values correspond to those of the projection method. As a result, except for this limiting case our test is strictly more powerful than the projection method and produces strictly smaller confidence sets in all realizations of the sample. Further, all of the assumptions we impose on the manifold S can be directly verified using the non-stochastic manifold $\widetilde{S}$ known to the researcher.

Concentrating out nuisance parameters. If one knows that the nuisance parameter β in a given testing problem is strongly identified then he/she can simply "concentrate out" the nuisance parameter, minimizing the AR statistic over β and reducing the degrees of freedom for the limiting distribution by k_β (that is, using quantiles of a $\chi^2_{k-k_\beta}$ rather than a χ^2_k). As discussed in section 2.1, this reduction in degrees of freedom stems from the fact that any manifold corresponding to a hypothesis with a strongly identified nuisance parameter converges to a linear subspace asymptotically.

The obvious advantage of this approach is that it is strictly more powerful than the projection method. However, the assumption of strong identification of the nuisance

parameter is essential, and the test may over-reject if this assumption fails. In many practical settings, including the DSGE and Phillips curve examples discussed above, the exact nature and source of weak identification is not clear, and we are unaware of any test of the null of weak identification which can be used to separate the weakly and strongly identified parameters. In contrast, the test we suggest in Section 3.4 does not employ any assumptions about the strength of identification of any parameter in point identified models. Indeed, since our approach is based on a finite-sample perspective (except perhaps for the assumed normality of the reduced-form parameters) we do not even require that there be a meaningful distinction between the weakly and strongly identified structural parameters in the model.

4.1 Pre-test for weak identification

If for some reason a researcher does not want to use our test, we suggest a simple procedure which could be called a "pre-test for weak identification." Imagine that a researcher wants to use a robust procedure (for example our test or the projection method) unless she knows that identification issues will not cause large size distortions, in which case she prefers instead to concentrate out the nuisance parameters. Our stochastic bounds can be used to address this question and determine whether weak identification constitutes a problem in a given setting. Below, we suggest a procedure which, when used as the first step of a two-step testing procedure of this sort, ensures that the procedure as a whole controls size.

To proceed, let us introduce the notion of a "tolerance level". Suppose that we would like to have a test of size α, but we are uncertain whether the usual strong-identification asymptotics provide a reasonable approximation in our context; in the event that these approximations are imperfect, we are willing to accept a test with true size $\alpha + \alpha^*$ in exchange for the additional power and convenience of using conventional critical values. The potential increase in the size α^* is called the tolerance level and has been used previously by e.g. Stock and Yogo (2005). For t-tests in weak IV, Stock and Yogo (2005) suggest comparing the minimal eigenvalue of the first-stage F statistic matrix to an appropriate threshold: if the minimal eigenvalue exceeds the threshold then the researcher can be confident that the usual 5% t-tests will have true size not exceeding 10%.

The pretest we propose asks whether the curvature of the model is sufficiently small to ensure that tests based on classical $\chi^2_{k-k_\beta}$ critical values (that is, tests which concentrate out the nuisance parameter β) with nominal size α have true size not exceeding $\alpha + \alpha^*$. To determine whether this is the case we calculate C^*, the smallest value C such that the $(1-\alpha)$-quantile of a χ^2_{k-p} distribution does not exceed $F_{1-(\alpha+\alpha^*)}(C, R, k, p)$. The cut-off C^* depends on the dimension k of the reduced-form parameter vector, the dimension p of the nuisance parameter, and R. To implement the pre-test we then calculate the value $\widehat{C}$ as described in Lemma 1, that is, the maximal curvature of the manifold at points inside a ball of radius $(1+\sqrt{2})R$ around the reduced-form estimator (noting that we may take $R = \infty$) and compare $\widehat{C}$ to C^*. If $\widehat{C} > C^*$ the researcher can safely concentrate out β and use $\chi^2_{k-k_\beta}$ critical values while if $\widehat{C} \leq C^*$ she should use a robust procedure. We can guarantee that the resulting two-step test will have size less than $\alpha + \alpha^*$.

Table 1 reports the cut-offs C^* for nominal 5% tests and tolerance level 5% for different values of p and k for R equal to the 0.99 quantile of a $\sqrt{\chi^2_k}$. Based on Table 1 we can see that for a fixed dimension k of the reduced-form parameter, increasing the number of nuisance parameters p tightens the restrictions imposed on curvature if one wants to concentrate out the nuisance parameters.

The main difference of this pre-test procedure from majority of existing tests of weak identification is that the resulting two-step procedure controls size. Tests of weak identification by Inoue and Rossi (2011), Iskrev (2010) and Wright (2003) all test the null of strong identification against the alternative of weak identification. Those tests control the probability of falsely rejecting strong identification but do not control the probability of failing to detect weak identification when it is present (this depends on power for these tests). As a result, using one of these tests in a two-step procedure, in which one tests the null of strong identification and uses a robust procedure only if strong identification is rejected, does not guarantee overall size control. In contrast, the procedure suggested here controls the size under weak identification.

4.2 AR conjecture

Kleibergen and Mavroeidis (2009a) consider a weak IV model with more than one endogenous regressor in which one wants to test a hypothesis about the coefficient on one endogenous regressor, treating the coefficients on the remaining regressors as nuisance

p\k	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.73	0.28	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2		2.33	1.94	1.65	1.41	1.20	1.00	0.82	0.65	0.49	0.34	0.19	0.05	0.00	0.00	0.00	0.00	0.00	0.00
3			4.00	3.63	3.36	3.13	2.93	2.74	2.57	2.41	2.26	2.11	1.97	1.84	1.71	1.59	1.46	1.35	1.23
4				5.72	5.38	5.11	4.88	4.68	4.50	4.34	4.18	4.03	3.89	3.76	3.63	3.50	3.38	3.27	3.15
5					7.46	7.16	6.90	6.68	6.47	6.29	6.13	5.97	5.83	5.69	5.55	5.43	5.31	5.19	5.07
6						9.23	8.96	8.71	8.49	8.29	8.10	7.94	7.78	7.63	7.49	7.36	7.24	7.11	7.00
7							11.00	10.78	10.55	10.32	10.12	9.94	9.77	9.61	9.46	9.32	9.19	9.06	8.94
8								12.80	12.62	12.40	12.18	11.98	11.79	11.62	11.46	11.31	11.17	11.03	10.90
9									14.58	14.46	14.25	14.05	13.85	13.66	13.48	13.32	13.17	13.03	12.89
10										16.38	16.31	16.12	15.92	15.72	15.54	15.37	15.20	15.05	14.90
11											18.18	18.16	18.00	17.80	17.61	17.43	17.25	17.08	16.93
12												19.97	20.02	19.88	19.69	19.51	19.32	19.15	18.98
13													21.78	21.88	21.76	21.59	21.41	21.22	21.05
14														23.58	23.75	23.65	23.49	23.31	23.14
15															25.38	25.62	25.55	25.40	25.23
16																27.19	27.49	27.45	27.31
17																	29.01	29.37	29.36
18																		30.83	31.25
19																			32.65

Table 1: Cut-off levels of C^* described in Section 4.1 for different values of k and p. Value of R is equal to 0.99- quantile of distribution χ^2_k. Based on 10 million simulations

parameters. They consider the AR statistic minimized over the nuisance parameters and show that its distribution under weak identification is dominated by its distribution under strong identification, and thus that χ^2 critical values with reduced degrees of freedom will always produce tests which maintain size. They also conjectured[5] that this statement could be generalized to *some* non-linear GMM models. If this statement held in all GMM models it would have eliminated the trade-off between projecting over and concentrating out nuisance parameters and implied that one should concentrate out in all cases to obtain more powerful tests while still controlling size. Below we provide a simple example which demonstrates that such a dominance result does not hold generally.

To fix ideas, assume that we have an i.i.d. sample $x_1, ..., x_n$ from a population that satisfies the k-dimensional moment condition $Ef_n(x_i, \theta) = 0$ at $\theta = \theta_0$ and assume that $\Sigma = \text{var}(f_n(x_i, \theta))$ is known. We first consider the AR statistic for testing the full parameter hypothesis $H_0 : \theta = \theta_0$:

$$AR(\theta_0) = \frac{1}{n}\left(\sum_{i=1}^{n} f_n(x_i, \theta_0)\right)\Sigma^{-1}\left(\sum_{i=1}^{n} f_n(x_i, \theta_0)\right)$$

This statistic is used with χ^2_k critical values. Assume that $\theta = (\alpha', \beta')'$ and we want to test the hypothesis $H_0 : \alpha = \alpha_0$ with nuisance parameter β. Consider the statistic:

$$AR(\alpha_0) = \min_{\beta} AR(\alpha_0, \beta).$$

We show that it is not in general true that the asymptotic distribution of $AR(\alpha_0)$ is stochastically dominated by a $\chi^2_{k-k_\beta}$, that is the limit distribution of $AR(\alpha_0)$ if β is strongly identified.

A counterexample to the AR conjecture. Consider an i.i.d. sample $x_1, ..., x_n$ drawn from $N(M_n(\theta_0), I_k)$ with k-dimensional structural parameter $\theta \in \Theta = [0, \pi]^{k-2} \times$

[5] The statement was made in one of the earlier versions of Kleibergen and Mavroeidis (2009a) and via private communication.

$[0, 2\pi) \times \mathbb{R} \subset \mathbb{R}^k$. Suppose further that

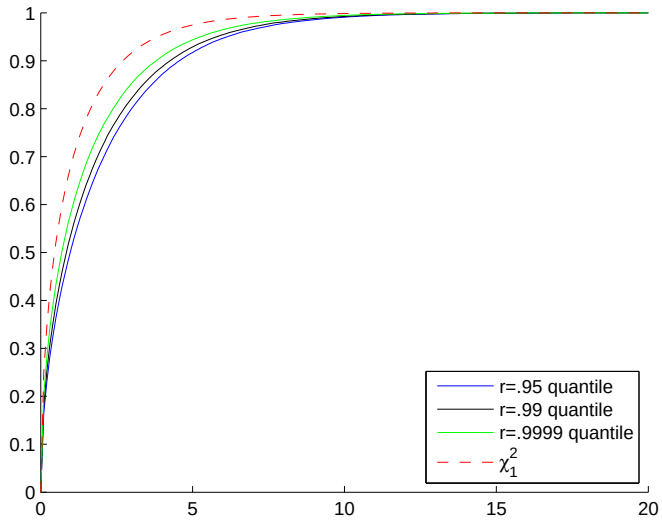
$$M_n(\theta) = \frac{r}{\sqrt{n}}\theta_k \begin{bmatrix} \cos(\theta_1) - 1 \\ \sin(\theta_1)\cos(\theta_2) \\ \sin(\theta_1)\sin(\theta_2)\cos(\theta_3) \\ \vdots \\ \sin(\theta_1)...\sin(\theta_{k-2})\cos(\theta_{k-1}) \\ \sin(\theta_1)...\sin(\theta_{k-2})\sin(\theta_{k-1}) \end{bmatrix}$$

is a k-dimensional vector-function of k variables. Note that all parameters are weakly identified in this case. The AR statistic for the full parameter vector is then

$$AR(\theta) = \frac{1}{n}\left(\sum_{i=1}^{n}\left[x_i - M_n(\theta)\right]\right)' \left(\sum_{i=1}^{n}\left[x_i - M_n(\theta)\right]\right).$$

Now assume that the hypothesis of interest is $H_0 : \theta_k = 1$, where $\beta = (\theta_1, ..., \theta_{k-1})'$ is a weakly identified nuisance parameter. The AR conjecture suggests that the limiting distribution of $\min_\beta AR(\beta, 1)$ is stochastically dominated by a χ^2_1. As before let us intro-

Figure 4: Distribution of statistic $\min_\beta AR(\beta, 1)$ for $k = 10$



duce a random variable $\xi = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[x_i - M_n(\theta_0)\right] \sim N(0, I_k)$, whose distribution does not depend on the sample size, and a function

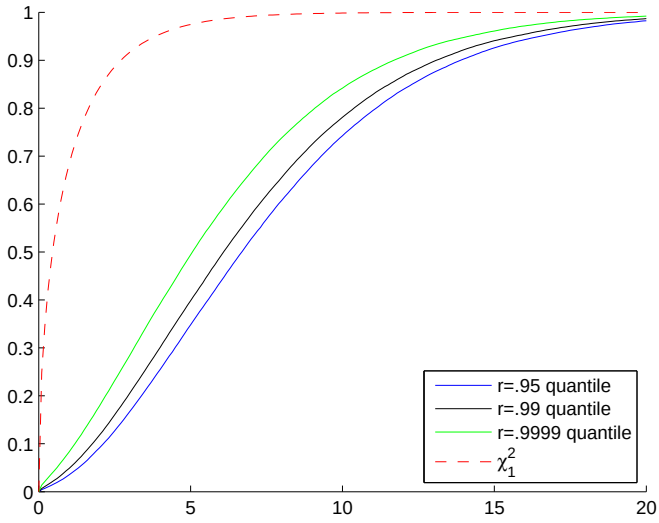
$$m(\beta) = \sqrt{n}\left(M_n(\beta, 1) - M_n(\theta_0)\right),$$

which also does not depend on the sample size. Notice that for all n the manifold described by the function $m(\beta)$ is a hyper-sphere of radius $r/\sqrt{n}$ in k-dimensional space passing through the origin. It is easy to see that

$$\min_{\beta} AR(\beta, 1) = \min_{\beta} \left(\xi - m(\beta)\right)' \left(\xi - m(\beta)\right).$$

Note, however, that the behavior of the last statistic does not depend on the sample size and can be easily simulated. Note further that since the distribution of a standard normal vector is rotation invariant, the distribution of $\min_\beta AR(\beta, 1)$ under the null depends only on k and the radius r. We simulate the distribution of the statistic of interest under the null for $k = 10$ and $k = 50$ and r equal to the square roots of the .95, .99, and .9999 quantiles of a χ^2_k distribution, and plot the resulting cdfs against a χ^2_1 cdf. Figures 4 and 5 show that tests which pair the minimized AR statistic with χ^2_1 critical values overreject. Further, the degree of over-rejection is increasing with the dimension k.

Figure 5: Distribution of statistic $\min_\beta AR(\beta, 1)$ for $k = 50$



4.3 Other methods

There are very few other papers that work directly with weakly identified nuisance parameters. One of them is D. Andrews and Cheng (2011). The authors impose some additional restrictions by assuming that they know the structure of weak identification, namely, they assume that it is known which parameters are potentially weakly identified

and that there is a parameter which defines the strength of identification. This assumption is not directly applicable to our DSGE and Phillips curve examples. Andrews and Cheng (2011) showed that in their case statistics for testing hypotheses with weakly identified nuisance parameters have non-standard asymptotic distributions which depend on the value of the nuisance parameter. Their procedure is based on simulating the asymptotic distribution of the test statistic for different values of the nuisance parameter and taking the "least favorable" among those distributions over a set of relevant nuisance parameter values.

Another example of inference with weakly identified nuisance parameters is given in I. Andrews and Mikusheva (2011). That paper considers a case when concentrating out a weakly identified parameter leads to asymptotically correct inferences, but this result holds only for weakly identified parameters which enter the log-likelihood function linearly.

5 Working with subset of parameters

Suppose we have a single observation $\widehat{\theta}$ from a population $\widehat{\theta} \sim N(\theta_0, \Sigma)$ with an unknown mean θ_0. We wish to test a hypothesis of the form $H_0 : \theta_0 = \theta(\beta)$ for some value of the p-dimensional structural parameter $\beta \in U \subset \mathbb{R}^p$. Our testing procedure suggested in Lemma 1 treats all components of the multi-dimensional vector β in such a way that only the direction of highest curvature affects the value of $\widehat{C}$ and thus influences the critical values. Imagine instead that β can be divided into two sub-sets of parameters $\beta = (\beta_1', \beta_2')'$ in such a way that the curvature corresponding to directions β_1 is high, but the null hypothesis seems to be close to flat in the parameter β_2. Let p_1 be the dimension of β_1, and p_2 the dimension of β_2: $p = p_1 + p_2$. In this section we propose modifications to our testing procedure (suggested in Lemma 1) and pre-test described in Section 4.1 that treat β_1 and β_2 differently. In particular we reduce the critical value of the test due to the low curvature with respect to β_2 while projecting over β_1. The modified procedure may be more efficient if the difference in curvature with respect to parameters β_1 and β_2 is large.

We start with a modification of the pre-test described in Section 4.1. This modification may be thought of as a pre-test for the possibility of concentrating out β_2 in the AR statistic defined in (14). The two competing options are 1) use our robust critical values

based on curvature for the full parameter vector or project over the full parameter vector; or 2) project over β_1, concentrate out β_2 and use $\chi^2_{k-p_2}$ critical values.

The suggested procedure is the following. For any value β_1 consider a p_2-dimensional manifold $S(\beta_1) = \{\Sigma^{-1/2} m(\beta_1, \beta_2), \beta_2 \in \mathbb{R}^{p_2}\}$. For any point $q = \Sigma^{-1/2} m(\beta_1, \beta_2) \in S(\beta_1)$ find the curvature $\kappa_q(S(\beta_1))$. Let

$$\widehat{C} = \left(\min_{\beta_1} \min_{q \in S(\beta_1) \bigcap \widetilde{B}} 1/\kappa_q(S(\beta_1)) \right) \wedge R, \tag{15}$$

where $\widetilde{B}$ is the ball of radius $(1+\sqrt{2})R$ around the point $\widehat{x} = \Sigma^{-1/2}\widehat{\theta}$. One also needs to check that as stated in Lemma 1 the analog of Assumption 1 is satisfied but now we are considering only p_2-dimensional balls in the tangent space to $S(\beta_1)$. If $\widehat{C} > C^*$ it is safe to use the second approach (that is, to concentrate out β_2), while otherwise the first approach should be used. We can likewise adapt the conclusion of Lemma 1 to state that the test which rejects the null if and only if

$$AR = \min_{\beta} (\widehat{\theta} - \theta(\beta))' \Sigma^{-1} (\widehat{\theta} - \theta(\beta)) > F_{1-\alpha}(\widehat{C}, R, k, p_2)$$

controls size.

Curvature of strongly identified parameters. If a set of nuisance parameters is strongly identified in the sense of Stock and Wright (2000), the null hypothesis is asymptotically linear in these parameters as we argued in Section 2.1. Here we show that the curvature corresponding to these parameters is asymptotically of order $O(1/\sqrt{n})$, where n is the sample size.

Consider a sample of size n from some model parameterized by structural parameter β that belongs to some bounded set $U \subseteq \mathbb{R}^p$ and assume that one can estimate the reduced-form parameters θ in a consistent and asymptotically normal way:

$$\sqrt{n}(\hat{\theta} - \theta) \Rightarrow N(0, \Sigma_0).$$

Assume that the relation between structural and reduced-form parameters $\theta = \theta(\beta)$ is fixed (not changing with n), twice continuously differentiable, and that the matrix $\frac{\partial}{\partial \beta} m(\beta)$ has full rank in a neighborhood of β_0, which is the only point in the closure of

U that solves the equation $\theta_0 = \theta(\beta)$. The null hypothesis manifold for sample size n is $S_n = \{\sqrt{n}\Sigma_0^{-1/2} m(\beta), \beta \in U\} \subset \mathbb{R}^k$. The maximal curvature over all points of the manifold S_n is equal to $1/\sqrt{n}$ multiplied by the maximal curvature of the manifold S_1 obtained for sample size 1. This can easily be seen by examining the role of the scale of m in formula (10).

6 Example: A Small-scale DSGE Model

To illustrate the utility of our theoretical results and suggest directions for future research, we apply our approach to a small-scale DSGE model based on Clarida, Gali and Gertler (1999). The (log-linearized) equilibrium conditions for the model are

$$\begin{cases} bE_t\pi_{t+1} + \kappa x_t - \pi_t + \varepsilon_t = 0, \\ -[r_t - E_t\pi_{t+1} - rr_t^*] + E_t x_{t+1} - x_t = 0, \\ \lambda r_{t-1} + (1-\lambda)\phi_\pi \pi_t + (1-\lambda)\phi_x x_t + u_t = r_t, \\ rr_t^* = \rho \Delta a_t, \end{cases} \tag{16}$$

where the exogenous variables (Δa_t and u_t) evolve according to

$$\Delta a_t = \rho \Delta a_{t-1} + \varepsilon_{a,t}; \quad u_t = \delta u_{t-1} + \varepsilon_{u,t};$$
$$(\varepsilon_t, \varepsilon_{a,t}, \varepsilon_{u,t})' \sim iid\, N(0, \Sigma); \Sigma = diag(\sigma^2, \sigma_a^2, \sigma_u^2).$$

Here we assume that a researcher observes data on inflation π_t, the interest rate r_t and some measure of real activity x_t. This model has ten parameters: the discount rate b, the structural parameters κ, ϕ_x, ϕ_π, and λ, and the parameters describing the evolution of the exogenous variables. We calibrate the structural parameters at generally accepted values similar to those used by Mueller (2010): $b = .99$, $\kappa = \frac{(1-\theta)(1+\phi)(1-b\theta)}{\theta} \approx .1717$, $\phi_x = 0.25$, $\phi_\pi = 1.5$ and $\lambda = 0.5$. For the parameters describing the exogenous variables, we choose $\rho = .2$ and $\delta = .2$ to introduce a degree of persistence while maintaining stationarity, and set $\sigma_a = .38$, $\sigma_u = .31$, and $\sigma = 1$. We generate samples of size 300 from this model and then discard the first 100 observations, using only the last 200 observations for the remainder of the analysis. Given well-documented problems with estimating b in many models, from this point forward we calibrate this parameter at its true value and treat the remaining 9 parameters as unknown.

In Andrews and Mikusheva (2011) we documented the poor performance of the classical Maximum Likelihood Estimator for this model and showed that classical confidence

sets are unreliable. In particular, we argued that the model displays behavior consistent with weak identification. Canova and Sala (2009) argued that many DSGE models suffer from weak identification. Simulation evidence in Ruge-Murcia (2010) also shows substantial overrejection for simulated method of moments-based Wald tests in some models. Unfortunately, however, the source and extent of weak identification in DSGE models is not well-understood. Most DSGE models are highly nonlinear in parameters, and are difficult if not impossible to solve analytically.

Several procedures have recently been proposed for testing simple hypotheses, that is hypotheses on the full parameter vector, in DSGE models which may be weakly identified. These tests include the score tests of I. Andrews and Mikusheva (2011) and Qu (2011), the LR statistic proposed by Guerron-Quintana, Inoue and Kilian (2009), and the test of Dufour, Khalaf, and Kichian (2009). Most of these papers suggest the projection method for testing hypotheses on subsets of parameters and confidence set construction, the exception being I. Andrews and Mikusheva (2011) who suggest a procedure for concentrating out strongly identified nuisance parameters. The main problem with the projection method is that it tends to be very conservative, since DSGE models typically have a large number of parameters. On the other hand, concentrating out nuisance parameters is also problematic as it is generally not clear which parameters are weakly identified, and we are unaware of any currently available procedure which would allow us to make this determination.

As outlined in section 2.2.1, to test hypotheses on the structural parameters in DSGE models we can test the implied restriction on the model auto-covariances $\theta(\beta)$. In particular, we let θ consist of the covariance matrix of the observables (x_t, π_t, r_t) and their first auto-covariance, giving us 15 reduced-form parameters. To focus on the problem of weak identification and abstract from the problems which may arise from HAC covariance matrix estimation, we treat the true covariance matrix Σ of our reduced-form parameter estimates as known, and consider AR statistics of the form $AR(\beta) = (\widehat{\theta} - \theta(\beta))' \Sigma^{-1} (\widehat{\theta} - \theta(\beta))$.

To illustrate the application of our approach, we consider the problem of separately testing that each of the structural parameters is equal to its true value (as one one needs to do to construct confidence sets for each parameter individually). For example, to test $H_0 : \kappa = \kappa_0$, we let $\widetilde{\beta}$ contain all the parameters other than κ and consider the AR statistic $AR(\kappa_0) = \min_{\widetilde{\beta}} AR(\kappa_0, \widetilde{\beta})$. As before, the key issue is what critical values to use. The projection method uses the 95th percentile of a χ^2_{15}, which is equal to 25. If

we assume that a q-dimensional sub-vector of $\widetilde{\beta}$ is strongly identified, we can use χ^2_{15-q} critical values instead, which are equal to 14.07 if we take all of $\widetilde{\beta}$ to be strongly identified ($q = 8$). Applying our robust critical values, in contrast, requires no assumption on the strength of identification. As we might expect in a poorly identified model, tests which concentrate out the nuisance parameters do not control size. We simulated tests for each parameter separately, and almost all of them over-reject, though the degree of over-rejection is limited. For example, nominal 5% tests for ρ and σ_a which concentrate out the nuisance parameters have size 9.2% and 9.6% respectively.

For each of the nine parameters, to compute our robust critical values we calculate the curvature of the submanifold of $\{\Sigma^{-\frac{1}{2}}\theta(\beta)\}$ obtained by holding that parameter equal to its null value, intersected with the ball $B_{\sqrt{2}R}(\theta(\beta_0))$ of radius $\sqrt{2}R$ around θ_0, for R the .99 quantile of a χ^2_{15} distribution. We find quite substantial curvature: the manifold implied by $\kappa = \kappa_0$, for example, has a maximal curvature of 1.72 which gives a robust critical value of 23.8. While smaller than the projection-method critical value, this is still quite large.

Not all parameters play an equal role in generating this curvature, however. As noted in I. Andrews and Mikusheva (2011), some parameters seem to be strongly identified, while other seem quite weak. To relate this to curvature, we consider projecting over different subsets of parameters as described in section 5. We find that by projecting over the Taylor-rule parameters ϕ_x and ϕ_π we can reduce the curvature dramatically, suggesting that the group of parameters other than ϕ_x and ϕ_π may be significantly better identified. Hence, to obtain smaller critical values, when testing hypotheses on the structural parameters we project over ϕ_x and ϕ_π.

For each structural parameter, Table 2 reports the robust critical value obtained from this exercise (column 2), together with the simulated size (based on 500 simulations) of nominal 5% tests based on our robust critical values (column 3), and projection-method tests (column 4). As we can see, projection-method based tests (using critical values of 25) are extremely conservative, with simulated size less than or equal to 0.2%. Our robust critical values range from 18.65 to 20.3 for different parameters, and the size of tests using these critical values (ranging from .4% to 1.6%) never exceeds the nominal size. At the same time, these tests are substantially less conservative than the projection method.

Parameter tested	Robust Critical Value	Robust Test Size	Projection Size
ϕ_x	18.72	1.20%	0.00%
ϕ_π	18.65	0.80%	0.00%
λ	19.86	1.60%	0.00%
ρ	19.68	1.40%	0.00%
δ	20.30	1.40%	0.20%
κ	19.51	1.40%	0.20%
σ_a	19.84	1.20%	0.00%
σ_u	19.27	0.40%	0.00%
σ	19.60	1.40%	0.00%

Table 2: Nominal 5% Tests of one-dimensional hypotheses on structural parameters. The first column lists the tested parameter for each row, while the other parameters are treated as nuisance parameters. The statistic is AR minimized over nuisance parameters. Projection method critical values are 25. Robust critical values are based on projecting over ϕ_x and ϕ_π.

7 Appendix with proofs

7.1 Proof of Theorem 1

The proof is based on the following lemma:

Lemma 2 *Assume the curve $\alpha(s) : [0, b] \to D_C \subset \mathbb{R}^k$ is parameterized by arc length and that its curvature $\kappa(s) = \|\ddot{\alpha}(s)\| < \frac{1}{C}$ for all points s. Assume that $\alpha(0) = 0$ and $\dot{\alpha}(0) = v \in$ span$\{e_1, ..., e_p\}$, where $e_1, ..., e_p$ are first p basis vectors. Then the curve $\alpha(s)$ is contained in the set $M_v \cap D_C$, where*

$$M_v = \{x : \langle x, v\rangle^2 + (C - \|x - \langle x, v\rangle v\|)^2 \geq C^2\}. \tag{17}$$

Proof of Lemma 2.

Consider the curve defined by $\beta(s) = \dot{\alpha}(s)$, the first derivative of α. Since the curve α is parameterized by arc length $\|\beta(s)\| = \|\dot{\alpha}(s)\| = 1$ and the new curve β lies on the unit sphere $Sph = \{x \in \mathbb{R}^k : \|x\| = 1\}$, with $\beta(0) = v$. Let $t \leq \frac{\pi}{2}C$ and $t \leq b$. Consider the arc length of the restriction of the curve β to the interval $[0, t]$:

$$length(t) = \int_0^t \|\dot{\beta}(s)\|ds = \int_0^t \|\ddot{\alpha}(s)\|ds = \int_0^t \kappa(s)ds \leq \frac{t}{C}.$$

This implies that the geodesic (a curve of a shortest length) on the sphere Sph connecting $\beta(0)$ and $\beta(t)$ has length less than or equal to $\frac{t}{C}$ or, equivalently, that the angle between

vectors $\beta(0) = v$ and $\beta(t)$ is less than or equal to $\frac{t}{C}$. Hence

$$\langle v, \beta(t)\rangle = \langle v, \dot{\alpha}(t)\rangle \geq \cos(\frac{t}{C}). \tag{18}$$

Since $\alpha(s)$ is parameterized by arc length, from inequality (18) we have:

$$\|\dot{\alpha}(t) - \langle v, \dot{\alpha}(t)\rangle v\| \leq |\sin(\frac{t}{C})|. \tag{19}$$

This, in turn, implies that

$$\|\alpha(t) - \langle v, \alpha(t)\rangle v\| = \|\int_0^t (\dot{\alpha}(s) - \langle v, \dot{\alpha}(s)\rangle v)ds\| \leq$$
$$\leq \int_0^t \|\dot{\alpha}(s) - \langle v, \dot{\alpha}(s)\rangle v\|ds \leq \int_0^t \sin(\frac{s}{C})ds = C - C\cos(\frac{t}{C})$$

Inequality (18) also implies that

$$\langle v, \alpha(t)\rangle \geq \int_0^t \cos(\frac{s}{C})ds = C\sin(\frac{t}{C}). \tag{20}$$

Combing these results yields

$$\langle v, \alpha(t)\rangle^2 + (C - \|\alpha(t) - \langle v, \alpha(t)\rangle v\|)^2 \geq C^2$$

for all $t \leq \frac{\pi}{2}C$. Notice that (20) implies that for $\tau = \frac{\pi}{2}C$ we have $\langle v, \alpha(\tau)\rangle \geq C$ and thus for the first p coordinates of $\alpha(\tau)$, which we denote $\alpha^{(1)}(\tau)$, we have $\|\alpha^{(1)}(\tau)\| \geq C$ so the curve is leaving or has already left the cylinder D_C and thus $b \leq \frac{\pi}{2}C$. This concludes the proof of the lemma. □

Proof of statement (a) of Theorem 1. First, let us show that

$$\bigcup_{\substack{v\in T_0(S)\\ \|v\|=1}} M_v = \{\|x^{(1)}\|^2 + (C - \|x^{(2)}\|)^2 \geq C^2\} = \mathcal{M}, \tag{21}$$

where M_v is defined in (17), $\mathcal{M}$ is defined in (11) and $T_0(S)$ is the tangent space to S at zero and is spanned by first p basis vectors. Indeed, the set on the left hand side consists

of points x for which there exists a vector $v \in span\{e_1, ..., e_p\}, \|v\| = 1$, such that

$$\langle x, v\rangle^2 + (C - \|x - \langle x, v\rangle v\|)^2 \geq C^2. \tag{22}$$

For each x let us find the maximum of the expression on left-hand side of inequality (22) over $v \in T_0(S), \|v\| = 1$:

$$\begin{aligned} &\langle x, v\rangle^2 + (C - \|x - \langle x, v\rangle v\|)^2 = \\ &= \langle x, v\rangle^2 + C^2 + \|x\|^2 - \langle x, v\rangle^2 - 2C\|x - \langle x, v\rangle v\| = \\ &= C^2 + \|x\|^2 - 2C\|x - \langle x, v\rangle v\| \end{aligned}$$

where we used that $\|x - \langle x, v\rangle v\|^2 = \|x\|^2 - \langle x, v\rangle^2$. We see that maximizing the left-hand side of (22) over $v \in span\{e_1, ..., e_p\}, \|v\| = 1$ is equivalent to minimizing $\|x - \langle x, v\rangle v\|$. The minimum is achieved at $v = \frac{1}{\|x^{(1)}\|}(x^{(1)}, 0, ..., 0)$, where $x^{(1)} \in \mathbb{R}^p$ consists of the first p components of x. As a result, the maximum of the left-hand side of (22) equals

$$C^2 + \|x\|^2 - 2C\|x^{(2)}\| = \|x^{(1)}\|^2 + (C - \|x^{(2)}\|)^2.$$

This proves statement (21).

Now assume that the statement (a) of Theorem 1 is incorrect and there exists a point $q \in S_C$ that $q \notin \mathcal{M}$. Take a geodesic line (a curve of the shortest distance lying in S_C) $\alpha(s)$ connecting q and 0 lying in S_C, where such curve exists since S_C is a connected manifold. Parameterize this curve by the arc length. The curve $\alpha(s)$ is geodesic in S if and only if at any point $q = \alpha(t)$ the second derivative $\ddot{\alpha}(t)$ is perpendicular to $T_q(S)$ (see Spivak (1999) for discussion of geodesics, v. 3, p.3). As a result, the curvature of the geodesic α at each point $q = \alpha(t)$ is equal to $\kappa_q(X, S)$ (where $X = \dot{\alpha}(t)$), and thus it is less than $\frac{1}{C}$. Denote the tangent to this curve at 0 by $v \in T_0(S)$. Applying Lemma 2 we obtain that the curve belongs to $M_v \cap D_C$ and thus belongs to $\mathcal{M} \bigcap D_C$. We have arrived at a contradiction. □

Proof of statement (c) of Theorem 1. Let $f(u) = \rho(\xi, N_u)$. We need to find the maximizer of $f(u)$ subject to the constraint $\|u\| = 1$. To differentiate $f(u)$ we use the "envelope theorem" that allows one to differentiate a function which is the optimum of a constrained optimization problem and yields $\frac{df(u)}{du} = \xi^{(2)} - zu$. Hence, the first-order

condition for finding $\widetilde{u}$ implies that u is proportional to $\xi^{(2)}$. The sign is a reflection of the fact that we search for a max rather than a min. $\square$

Proof of statement (b) of Theorem 1. For a given point $\xi \in \mathbb{R}^k$ find the sphere $N_{\widetilde{u}}$ furthest from ξ, $\widetilde{u}$ is described in Theorem 1 (c), and the point $\tau \in N_{\widetilde{u}}$ such that $\rho(\xi, N_{\widetilde{u}}) = \rho(\xi, \tau)$. Consider the $k-p$ dimensional linear space $R_\tau = \{x \in \mathbb{R}^k : x^{(1)} = \tau^{(1)}\}$ that restricts the first p components of x to coincide with the first p components of τ. We will put forward two statements: first, that all points in the intersection of $R_\tau \bigcap \mathcal{M} \bigcap D_C$ are not further from ξ than τ; and second, that this intersection $R_\tau \bigcap \mathcal{M} \bigcap D_C$ contains at least one point from S. Together, these two statements imply that $\rho(\xi, S) \leq \rho(\xi, \tau)$.

The intersection of the three sets $R_\tau \bigcap \mathcal{M} \bigcap D_C$ can be written as follows:

$$R_\tau \bigcap \mathcal{M} \bigcap D_C = \{x = (\tau^{(1)}, x^{(2)}) \in D_C : \|\tau^{(1)}\|^2 + (C - \|x^{(2)}\|)^2 \geq C^2\} = \\ = \left\{x = (\tau^{(1)}, x^{(2)}) : \|x^{(2)}\| \leq C - \sqrt{C^2 - \|\tau^{(1)}\|^2}\right\}.$$

Now let us show that for each $x \in R_\tau \bigcap \mathcal{M} \bigcap D_C$ we have $\rho(\xi, x) \leq \rho(\xi, \tau)$. Indeed, one can solve the constrained maximization problem

$$\rho(\xi, x)^2 = \|\xi^{(1)} - \tau^{(1)}\|^2 + \|\xi^{(2)} - x^{(2)}\|^2 \to \max \text{ s.t. } x \in R_\tau \bigcap \mathcal{M} \bigcap D_C.$$

From the first-order condition for this problem one can see that the maximum is achieved at $x^{(2)}$ proportional to $\xi^{(2)}$, and further inspection reveals that it is achieved at $x = \tau$. Hence, all points lying in the intersection $R_\tau \bigcap \mathcal{M} \bigcap D_C$ have distance to ξ less or equal than $\rho(\xi, N_{\widetilde{u}})$.

To complete the proof we need only show that $R_\tau \bigcap \mathcal{M} \bigcap D_C$ contains at least one point from the manifold S. Recall that from the definition of $\tau \in N_{\widetilde{u}}$ it follows that $\|\tau^{(1)}\| \leq C$. Then Assumption 1 guarantees that the intersection of S_C with R_τ is non-empty, while statement (a) of Theorem 1 implies that $S_C \subseteq \mathcal{M} \bigcap D_C$. $\square$

Proof of statement (d) of Theorem 1. Note that both ξ and $N_{\widetilde{u}}$ belong to the same $p+1$- dimensional linear sub-space $L_{\widetilde{u}} = \{x : x = (x^{(1)}, -z\widetilde{u}), x^{(1)} \in \mathbb{R}^p, z \in \mathbb{R}\}$. Let us restrict our attention to this subspace only. Let $(x^{(1)}, z)$ be the coordinate system in this sub-space, so ξ corresponds to $\widetilde{\xi} = (\xi^{(1)}, \|\xi^{(2)}\|)$, and $N_{\widetilde{u}}$ corresponds to the sphere $N^C = \{x = (x^{(1)}, z) \in \mathbb{R}^{p+1} : \|x^{(1)}\|^2 + (C + z)^2 = C^2\}$. The distance implied by the

distance in $\mathbb{R}^k$ is the usual Euclidean metric, which we denote by $\widetilde{\rho}$. So far, we proved that $\rho(\xi, N_{\widetilde{u}}) = \widetilde{\rho}(\widetilde{\xi}, N^C)$. By invariance of the distance to orthonormal transformations of first p components we have $\widetilde{\rho}(\widetilde{\xi}, N^C) = \widetilde{\rho}(\xi^*, N^C)$, where $\xi^* = (\|\xi^{(1)}\|, 0, ..., 0, \|\xi^{(2)}\|) \in \mathbb{R}^{p+1}$. From this it is easy to see that

$$\rho(\xi, N_{\widetilde{u}}) = \rho_2(\eta, N_2^C),$$

where $\eta = (\|\xi^{(1)}\|, \|\xi^{(2)}\|) \in \mathbb{R}^2$, $N_2^C = \{(z_1, z_2) \in \mathbb{R}^2 : z_1^2 + (C + z_2)^2 = C^2\}$, and ρ_2 is Euclidian distance in $\mathbb{R}^2$. It then follows that if $\xi \sim N(0, I_k)$ then components of η have independent $\sqrt{\chi_p^2}$ and $\sqrt{\chi_{k-p}^2}$ distributions, respectively. $\square$

7.2 Proof of Lemma 1

Proof of Lemma 1. Let $\xi = \Sigma^{-1/2}(\widehat{\theta} - \theta_0) \sim N(0, I_k)$ and $S = \{\Sigma^{-1/2}(\theta - \theta_0), \theta \in H_0\} \subset \mathbb{R}^k$. Let $\psi_C(\xi, R)$ be defined as

$$\psi_C(\xi, R) = \begin{cases} \rho^2(\xi, N_{\widetilde{u}}), & \text{if } \|\xi\| \leq R; \\ \|\xi\|^2, & \text{if } \|\xi\| > R, \end{cases}$$

where $N_{\widetilde{u}} = \{x \in \mathbb{R}^k : x = (x^{(1)}, z\widetilde{u}), x^{(1)} \in \mathbb{R}^p, z \in \mathbb{R}_+, \|x^{(1)}\|^2 + (C - z)^2 = C^2\}, \widetilde{u} = -\frac{1}{\|\xi^{(2)}\|}\xi^{(2)}$. Consider the infeasible test φ which rejects ($\varphi = 1$) if and only if $\psi_C(\xi, R) \geq F_{1-\alpha}(C, R, k, p)$. The size $E\varphi(\xi) = \alpha$, so since $P\{\chi_k^2 \geq R^2\} < \alpha$ we know that φ rejects for all realizations of ξ where $\|\xi\| > R$. This test is infeasible, however, since we do not know the true value of θ_0 and hence cannot calculate ξ. The (feasible) test described in Lemma 1 is

$$\widetilde{\varphi} = \begin{cases} 1, & \text{if } AR \geq F_{1-\alpha}(\widehat{C}, R, k, p); \\ 0, & \text{otherwise.} \end{cases} \tag{23}$$

We claim that $\widetilde{\varphi} \leq \varphi$ almost surely (realization-by-realization). To show that this is the case, assume that $\widetilde{\varphi} = 1$. If at the same time $\|\xi\| > R$ then $\varphi = 1$, so the claim holds. If, on the other hand, $\|\xi\| \leq R$, then the cylinder $\widetilde{D}_R(x_0)$ around $x_0 = \Sigma^{-1/2}\theta_0$ lies inside of ball $\widetilde{B}$, and thus

$$\widehat{C} = \left(\min_{q \in \widetilde{S} \bigcap \widetilde{B}} 1/\kappa_q(\widetilde{S}) \right) \wedge R \leq \left(\min_{q \in \widetilde{S} \bigcap \widetilde{D}_R(x_0)} 1/\kappa_q(\widetilde{S}) \right) \wedge R \leq C.$$

Indeed, to justify the last inequality, consider two cases $R \leq C$ and $R > C$. In the first case $\widehat{C} \leq R \leq C$, in the second case $\widetilde{D}_C \subset \widetilde{D}_R$ and thus $\min_{q \in \widetilde{S} \bigcap \widetilde{D}_R(x_0)} 1/\kappa_q(\widetilde{S}) \leq \min_{q \in \widetilde{S} \bigcap \widetilde{D}_C(x_0)} 1/\kappa_q(\widetilde{S}) \leq C$.

Note that the function $F_{1-\alpha}(c, R, k, p)$ is decreasing in c, and hence $F_{1-\alpha}(C, R, k, p) \leq F_{1-\alpha}(\widehat{C}, R, k, p)$. Further, all the assumptions of Theorem 1 are satisfied so $AR = \rho^2(\xi, S) \leq \rho^2(\xi, N_{\widetilde{u}}) \leq \psi_C(\xi, R)$. Combining these results we obtain that

$$F_{1-\alpha}(C, R, k, p) \leq F_{1-\alpha}(\widehat{C}, R, k, p) \leq AR = \rho^2(\xi, S) \leq \psi_C(\xi, R),$$

and thus $\varphi = 1$. Hence whenever $\widetilde{\varphi} = 1$, we get that $\varphi = 1$ as well, so $\widetilde{\varphi} \leq \varphi$ as we wanted to show, and the size of the feasible test $\widetilde{\varphi}$ is bounded above by α, completing the proof. □

8 References

Anderson, G., and G. Moore (1985): "A Linear Algebraic Procedure for Solving Linear Perfect Foresight Models," *Economics Letters*, 17(3),247-252.

Andrews, D.W.K., and X. Cheng (2011): "Estimation and Inference with Weak, Semi-strong and Strong Identification," *Econometrica, forthcoming.*

Andrews, I. and A. Mikusheva (2011): "Maximum Likelihood Inference in Weakly Identified Models," *unpublished manuscript.*

Blanchard, O.J., and C.M. Kahn (1980): "The Solution of Linear Difference Models under Rational Expectations," *Econometrica*, 48(5),1305-1311.

Canova, F., and L. Sala (2009): "Back to Square One: Identification Issues in DSGE Models," *Journal of Monetary Economics*, 56, 431-449.

Christiano L.J. and M. Eichenbaum (1992): "Current Real-Business-Cycle Theories and Aggregate Labor-Market Fluctuations," *American Economic Review*, 82, 430-450.

Clarida, R., J. Gali, and M. Gertler (1999): "The Science of Monetary Policy: A New Keynesian Perspective," *Journal of Economic Literature*, 37, 1661-1707.

Dufour, J.M., (2003): "Identification, Weak Instruments, and Statistical Inference in Econometrics," *The Canadian Journal of Economics*, 36, 767-808.

Dufour, J.-M., Jasiak, J., (2001): "Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors," *International Economic*

Review 42, 815-843.

Dufour, J.M., L. Khalaf, and M. Kichian (2006): "Inflation Dynamics and the New Keynesian Phillips Curve: An Identification Robust Econometric Analysis," *Journal of Economic Dynamic and Control*, 30, 1707-1727.

Dufour, J.M., L. Khalaf, and M. Kichian (2009): "Structural Multi-Equation Macroeconomic Models: Identification-Robust Estimation and Fit," *Bank of Canada Working Paper*.

Dufour, J.M., and M. Taamouti (2005): "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," *Econometrica*, 73, 1351-1365.

Efron, B. (1975): "Defining the Curvature of a Statistical Problem (with Applications to Second Order Efficiency)," *Annals of Statistics*, 3, 1189-1242.

Gali, J., and M. Gertler (1999): "Inflation Dynamics: A Structural Econometric Analysis," *Journal of Monetary Economics* 44, 195-222.

Guerron-Quintana, P., A. Inoue, and L. Kilian (2009): "Frequentist Inference in Weakly Identified DSGE Models," *unpublished manuscript*.

Guggenberger, P. and R.J. Smith (2005): "Generalized Empirical Likelihood Estimators and Tests under Partial, Weak and Strong Identification," *Econometric Theory*, 21, 667-709.

Guggenberger, P. and R.J. Smith (2008): "Generalized Empirical Likelihood Tests in Time Series Models with Potential Identification Failure," *Journal of Econometrics*, 142(1), 134-161.

Inoue, A. and B. Rossi (2011): "Testing for Weak Identification in Possibly Nonlinear Models," *Journal of Econometrics*, 161(2), 246-261.

Iskrev, N. (2010): "Local Identification in DSGE," *Journal of Monetary Economics*, 57(2), 189-202.

King, R.G. and M.W. Watson (1998): "The Solution of Singular Linear Difference Systems under Rational Expectations," *International Economic Review*, 39(4), 1015-1026.

Kleibergen, F. (2005): "Testing Parameters in GMM Without Assuming that They Are Identified," *Econometrica*, 73, 1103-1123.

Kleibergen, F. (2007): "Generalizing Weak Instrument Robust IV Statistics towards Multiple Parameters, Unrestricted Covariance Matrices and Identification Statistics," *Journal of Econometrics*, 139(1), 181-216.

Kleibergen, F., and S. Mavroeidis (2009a): "Inference on Subsets of Parameters in GMM without Assuming Identification," *unpublished manuscript*.

Kleibergen, F., and S. Mavroeidis (2009b): "Weak Instrument Robust Tests in GMM and

the New Keynesian Phillips Curve," *Journal of Business & Economic Statistics*, 27(3), 293-311.

Kobayashi, S. and K. Nomizu (1969): "*Foundations of Differential Geometry*," *Interscience Publishers.*

Mavroeidis, S. (2005): "Identification Issues in Forward-looking Models Estimated by GMM with an Application to the Phillips Curve," *Journal of Money, Credit and Banking*, 37, 421-449.

Mavroeidis, S. (2010): "Monetary Policy Rules and Macroeconomic Stability: Some New Evidence," *American Economic Review*, 100(1), 491-503.

Mueller, U. (2010): "Measuring Prior Sensitivity and Prior Informativeness in Large Bayesian Models," *unpublished manuscript.*

Nason, J.M. and G.W. Smith (2008): "Identifying the New Keynesian Phillips Curve," *Journal of Applied Econometrics*, 23, 525-551.

Qu, Z. (2011): "Inference and Specification Testing in DSGE Models with Possible Weak Identification," *unpublished manuscript.*

Rotemberg, J., and M. Woodford (1997): "An Optimization Based Econometric Framework for the Evaluation of Monetary Policy," *NBER Macroeconomic Annual*, 12, 297-346.

Ruge-Murcia, F.J. (2007): "Methods to Estimate Dynamic Stochastic General Equilibrium Models," *Journal of Economic Dynamics & Control*, 31, 2599-2636.

Ruge-Murcia, F.J. (2010): "Estimating Nonlinear DSGE Models by the Simulated Method of Moments: With an Application to Business Cycles," *unpublished manuscript.*

Sims, C.A. (2002): "Solving linear rational expectations models," *Computational Economics*, 20, 1-20.

Spivak, M. (1999): "*A Comprehensive Introduction to Differential Geometry*," *Publish or Perish, Texas.*

Stock, J. H., and J.H. Wright (2000): "GMM With Weak Identification," *Econometrica*, 68, 1055-1096.

Stock, J.H., J.H. Wright, and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics*, 20(4), 518-529.

Stock, J.H. and M. Yogo (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg,* D. W. K. Andrews and J. H. Stock, eds., Cambridge: Cambridge University

Press, pp. 80-108.

Wright, J.H. (2003): "Detecting Lack of Identification in GMM," *Econometric Theory*, 19, 322-330.

Yogo, M., (2004): "Estimating the Elasticity of Intertemporal Substitution when Instruments are Weak," *Review of Economics and Statistics* 86, 797-810.

A Geometric Approach to Weakly Identified Econometric Models

By Isaiah Andrews[1] and Anna Mikusheva[2]

Abstract

Many nonlinear Econometric models show evidence of weak identification, including many Dynamic Stochastic General Equilibrium models, New Keynesian Phillips curve models, and models with forward-looking expectations. In this paper we consider minimum distance statistics and show that in a broad class of models the problem of testing under weak identification is closely related to the problem of testing a "curved null" in a finite-sample Gaussian model. Using the curvature of the model, we develop new finite-sample bounds on the distribution of Anderson-Rubin-type statistics, which we show can be used to detect weak identification and to construct tests robust to weak identification. We apply the new method to a small-scale DSGE model and show that it provides a significant improvement over existing methods.

Key words: weak identification, statistical differential geometry

This draft: May 29, 2012.

1 Introduction

Empirical researchers in Economics frequently find that even in large samples the data provides little information about some model parameters. In such cases, known as weakly identified, the usual asymptotic approximations for estimators and test statistics may be quite poor, making standard approaches to inference unreliable. Weak identification has been detected in a wide range of *non-linear* estimation contexts, including estimation of the New Keynsian Phillips Curve (Dufour, Khalaf, and Kichian (2006), Kleibergen and Mavroeidis (2009b), Mavroeidis (2005), Nason and Smith (2008)), monetary policy rules (Mavroeidis (2010)), Dynamic Stochastic General Equilibrium (DSGE) Models (Ruge-Murcia (2007), Canova and Sala (2009), Iskrev (2010), I. Andrews and Mikusheva (2011), Guerron-Quintana, Inoue and Kilian (2009)), and Euler equations (Yogo (2004)).

[1]Department of Economics, M.I.T., 50 Memorial Drive, Building E52, Cambridge, MA, 02142. Email: iandrews@mit.edu. Financial support from the NSF Graduate Research Fellowship Program is gratefully acknowledged.

[2]Department of Economics, M.I.T., 50 Memorial Drive, Building E52, Cambridge, MA, 02142. Email: amikushe@mit.edu. Financial support from the Castle-Krob Career Development Chair is gratefully acknowledged. We are extremely grateful to Professor of Differential Geometry Toby Colding for discussing geometric issues with us, and for checking our geometric proofs. We are grateful to Gary Chamberlain, Victor Chernozhukov, Jerry Hausman, Mattias Cattaneo, Whitney Newey and Jim Stock for helpful comments.

The need for more reliable procedures robust to weak identification in non-linear contexts has inspired a large literature in econometrics - for a survey, see Dufour (2003) and Stock, Wright, and Yogo (2002).

A number of different testing procedures have been proposed in this literature, most of which address two situations: the case in which one is interested in testing the full parameter vector (also known as testing a simple hypothesis), and the case in which one is interested in testing only a subset of parameters but the parameters not under test (the nuisance parameters) are strongly identified. Examples of such tests include Stock and Wright (2000), Guggenberger and Smith (2005, 2008), Kleibergen (2005, 2007), I. Andrews and Mikusheva (2011), and Qu (2011). The literature to date has, however, been largely silent about the case in which part of the nuisance parameter vector may be weakly identified. A notable exception is the recent paper by D. Andrews and Cheng (2011).

Our paper directly addresses the question of inference with weakly identified nuisance parameters in the context of minimum distance estimation. We suggest a fully robust testing procedure which controls size without any assumption on the strength of identification of the parameters. Further, if the nuisance parameters are strongly identified, our procedure is asymptotically equivalent to the "concentrated out" S-test suggested by Stock and Wright (2000) for hypotheses with strongly identified nuisance parameters.

Our procedure is based on novel finite-sample bounds on the distribution of the test statistic under the null. For our asymptotics, we assume only that the model has reduced-form parameters which are asymptotically normal and hence we do not rely on any particular asymptotic embedding, such as those used by Stock and Wright (2000) or D. Andrews and Cheng (2011) to model weak identification. In fact, if one thinks a normal approximation to the distribution of the reduced-form estimates is reasonable, our bounds are finite-sample rather than asymptotic.

The bounds we derive rely on techniques from differential geometry which are new in the econometrics literature. Our starting point is the observation that hypotheses in non-linear models with strongly identified nuisance parameters are asymptotically linear in a geometrical sense. In contrast, hypotheses with weakly identified nuisance parameters need not be asymptotically linear and can exhibit substantial curvature even in large samples, leading to the breakdown of the usual asymptotic approximations. Our bounds can be viewed as a strengthening of the usual approximations, where rather

than appealing to asymptotic linearity of the null hypothesis we quantify the maximal deviation of the null from linearity and use it to construct stochastic bounds. As a result, under strong-identification asymptotics our bounds recover the usual approximations. The test we suggest uses a standard minimum-distance statistic paired with robust critical values derived using these geometric bounds. The validity of this approach relies only on assumptions which can be checked directly in applications, and our robust critical values are very easy to simulate. The bounds we derive are also of potential interest for a range of other applications, including testing nonlinear hypotheses and inference in highly non-linear models. Our approach differs from the statistical geometry literature initiated by Efron (1975) in that we produce finite sample bounds on the distribution of the test statistic, whereas the statistical geometry literature is primarily concerned with higher-order asymptotic approximations.

To date the dominant recommendation for testing hypotheses with weakly identified nuisance parameters has been the projection method (see Dufour and Jasiak (2001), Dufour and Taamouti (2005), Dufour, Khalaf, and Kichian (2006)). The strength of the projection method is that it requires no assumptions beyond the validity of the test for the full parameter vector. It is in general conservative, however, and may be extremely so in cases where the nuisance parameter is high-dimensional and/or strongly identified. Our approach is an improvement over the projection method, in that it uses the same test statistic paired with smaller critical values while still maintaining size.

If one knows that part of the nuisance parameter vector is strongly identified, it has been proved that in many cases one can obtain a more powerful test by concentrating out the nuisance parameter as in e.g. Stock and Wright (2000). Maintaining correct size in such cases, however, relies critically on the strong identification assumption on the nuisance parameter. In contrast, our approach requires no assumption of strong identification but, in the event that the nuisance parameters are strongly identified, is asymptotically equivalent to concentrating them out. In this sense, our robust critical values can be viewed as providing a continuous transition between projecting over and concentrating out the nuisance parameters, depending on the strength of identification.

We also use our bounds to derive a pre-test for weak identification which allows one to control size in two-step testing procedures. In particular, if one is deciding between concentrating out the nuisance parameter and using a robust procedure, we provide bounds on the curvature of the null hypothesis which can be used to guide the choice.

This procedure is quite different from existing tests for weak identification in nonlinear models such as Inoue and Rossi (2011), Iskrev (2010) and Wright (2003) in that the existing tests all effectively test the null of strong identification and hence do not control the probability of failing to detect weak identification. In contrast, our pre-test for weak identification directly bounds the finite-sample distribution of the test statistic.

As a side result, we provide an example showing that the distribution of an AR-type statistic in a nonlinear GMM model under weak identification is *not* in general dominated by the distribution of the same statistic under strong identification. Kleibergen and Mavroeidis (2009a: initial draft) claimed that such such stochastic dominance holds for some non-linear GMM models. If such a statement held universally (which Kleibergen and Mavroeidis did not claim), it would have meant that concentrating out the nuisance parameter controlled size regardless of identification strength. As we show in this paper, however, this dominance relationship does not in general hold for our setting, and concentrating out weakly identified nuisance parameters can yield substantial over-rejection.

We apply our approach to a small-scale DSGE model and find evidence of substantial curvature. We consider the problem of testing composite hypotheses about model parameters, and show that our robust critical values are substantially smaller than those used by the projection method while still controlling size.

The paper is structured as follows. In Section 2 we show that hypotheses with strongly identified nuisance parameters are asymptotically linear, while weakly identified nuisance parameters may cause non-trivial curvature of the null hypothesis. We also introduce several examples that fit our framework. In Section 3 we derive our geometric and stochastic bounds and introduce our fully robust test. In Section 4 we compare our testing procedures with existing methods, discuss the AR conjecture and introduce our pre-test. Section 5 is devoted to modified procedures for subsets of nuisance parameters, and Section 6 presents simulation results from applying our procedures to a small-scale DSGE model. All proofs may be found in the Appendix.

Throughout the paper we use the following notation: $\dot{\alpha}$ is the derivative of the function α, $\ddot{\alpha}$ is the second derivative, $B_R(x_0) = \{x \in \mathbb{R}^k : \|x - x_0\| \leq R\}$ is a k-dimensional ball of radius R with center x_0, and $B_R = B_R(0)$ is a ball around zero. Let $D_C = \{x = (x^{(1)}, x^{(2)}) : \|x^{(1)}\| \leq C, \|x^{(2)}\| \leq C, x^{(1)} \in \mathbb{R}^p, x^{(2)} \in \mathbb{R}^{k-p}\} \subset \mathbb{R}^k$, which is a natural generalization of a cylinder.

2 Model setting

Suppose we have a single observation $\hat{\theta} \sim N(\theta, I_k)$, where the true value of $\theta = \theta_0$ is unknown. Further, suppose we have a known regular p-dimensional manifold $\widetilde{S}$ in k-dimensional space. The assumption that $\widetilde{S}$ is a regular p-dimensional manifold essentially means that the tangent space to $\widetilde{S}$ at all points is a p-dimensional linear subspace.[3] The LR (likelihood ratio) statistic for testing a null hypothesis $H_0 : \theta \in \widetilde{S}$ is

$$LR = \min_{\theta \in \widetilde{S}} (\hat{\theta} - \theta)'(\hat{\theta} - \theta).$$

Example 1. Let $g : \mathbb{R}^k \to \mathbb{R}^{k-p}$ be a twice-continuously-differentiable function whose Jacobian has full rank at all points. Assume that we wish to test the null hypothesis $H_0 : g(\theta) = 0$. Then the set of points $\widetilde{S} = \{\theta \in \mathbb{R}^k : g(\theta) = 0\}$ describes a p-dimensional manifold which is known to the researcher. If g is a linear function, the null can be formulated as $H_0 : R\theta = r$, where R is $(k-p) \times k$ full-rank matrix and r is $(k-p) \times 1$ vector. In this special case, the manifold $\widetilde{S}$ described by the null hypothesis is a p-dimensional linear space.□

Example 2. Suppose we have a structural model which imposes that $\theta = \widetilde{m}(\beta)$ for some structural parameter β where the function $\widetilde{m} : \mathbb{R}^p \to \mathbb{R}^k$ is twice continuously differentiable with a full-rank Jacobian at all points. The image of the function $\widetilde{m}$ is a p-dimensional regular manifold $\widetilde{S}$ known to the researcher. Hence, the LR test for the hypothesis of correct specification will be based on the statistic

$$LR = \min_{\beta} (\hat{\theta} - \widetilde{m}(\beta))'(\hat{\theta} - \widetilde{m}(\beta)) = \min_{\theta \in \widetilde{S}} (\hat{\theta} - \theta)'(\hat{\theta} - \theta). \ \square$$

The distribution of the LR statistic under the null, that is for $\theta \in \widetilde{S}$, is in general non-standard and depends on the unknown nuisance parameter θ_0, the true value of θ, making inference difficult. In what follows, we derive bounds on distribution of LR for which it suffices to know the shape of the manifold $\widetilde{S}$; that is, which do not require knowledge of θ_0.

To proceed, it is useful for us to introduce the random vector $\xi = \hat{\theta} - \theta_0 \sim N(0, I_k)$ and the p-dimensional manifold $S = \{x : x = \theta - \theta_0, \theta \in \widetilde{S}\}$, which is simply $\widetilde{S}$ translated by $-\theta_0$. If the null is true, then the manifold S passes through the origin. The LR

[3] More details on regularity conditions can be found in section 3.

statistic is equal to the squared distance between ξ and S:

$$\rho(\xi, S)^2 = \min_{x \in S} (\xi - x)'(\xi - x). \tag{1}$$

The central statistical issue in this paper is how to characterize the distribution of $\rho(\xi, S)$, the squared distance from ξ to the manifold S, in terms which do not depend on the unknown θ_0. In particular, we develop bounds on this distribution that depend only on the curvature of the manifold $\widetilde{S}$, which is known to the researcher.

A well-known property of the normal distribution is that if S is a p-dimensional linear sub-space then the squared distance $\rho^2(\xi, S)$ has a χ^2_{k-p} distribution. Note that in this very special case the distribution does not depend on θ_0. Numerous classical results on testing in the presence of a nuisance parameter are based on this fact. Indeed, most of the classical statistics literature deals with testing hypotheses that are either linear or asymptotically linear, in the sense that S is either a linear subspace or arbitrarily well-approximated by one in large samples. In subsection 2.1 below, we argue that testing in the presence of strongly identified nuisance parameters is asymptotically equivalent to testing a linear hypothesis, while testing in the presence of weakly identified nuisance parameters tends to result in asymptotically non-linear null hypotheses. This has important implications for hypothesis testing, since if S is not a linear sub-space then the distribution of $\rho^2(\xi, S)$ is in general non-standard and depends on the whole shape of S.

There is one natural (less informative) bound that can be placed on $\rho^2(\xi, S)$ without any assumptions, namely that $\rho^2(\xi, S)$ is dominated by χ^2_k. Indeed, since $0 \in S$, we have

$$\rho(\xi, S)^2 = \min_{x \in S} (\xi - x)'(\xi - x) \leq (\xi - 0)'(\xi - 0) \sim \chi^2_k. \tag{2}$$

We argue below that this bound is precisely the one used by the "projection method", which is currently the main approach available for testing with weakly identified nuisance parameters. As discussed above, the major disadvantage of this bound is that it may yield quite conservative tests.

2.1 Weak identification is related to curvature

A parameter or group of parameters is known as weakly identified when it is point identified, but the data is not very informative about the true value. In such cases, it is

well known that many classical statistical approximations perform quite poorly. As a result, point estimates tend to be biased, many tests exhibit large size distortions, and confidence sets based on these tests have poor coverage. One framework used for developing better approximations in such cases is the drifting functions approach introduced by Stock and Wright (2000). This approach models weak identification using an asymptotic embedding in which the objective function is asymptotically flat along some directions. To fix ideas, let us consider a GMM model in which the moment function is separable in the data. In particular, assume that we observe a sample $\{x_i\}$ of size n consisting of identically and independently distributed observations such that

$$E(h(x_i) - M(\alpha, \beta)) = 0 \quad \text{for} \quad \alpha = \alpha_0, \ \beta = \beta_0. \tag{3}$$

Here $h(x)$ is a k-dimensional function with $E\|h(x_i)\|^4 < \infty$, while α and β are $k_\alpha \times 1$ and $k_\beta \times 1$ vectors respectively, for $k_\alpha + k_\beta \leq k$. Assume that $\theta_0 = (\alpha_0, \beta_0)$ is the unique point at which the moment condition (3) is satisfied, so that the model is point identified. As in Stock and Wright (2000), we can allow the function M to change as the sample size grows. In particular,

$$M(\alpha, \beta) = M_n(\alpha, \beta) = \widetilde{M}(\alpha) + \frac{1}{\sqrt{n}} M^*(\alpha, \beta), \tag{4}$$

where $\widetilde{M}(\alpha)$ and $M^*(\alpha, \beta)$ are fixed twice-continuously-differentiable functions with full-rank Jacobians. In this setting, α is strongly identified while β is weakly identified, because information about β does not accumulate as the sample size grows.

Suppose we are interested in testing a hypothesis about the structural parameters α and β. Consider first the problem of testing a full parameter vector hypothesis $H_0 : \alpha = \alpha_0, \beta = \beta_0$. To test this hypothesis, we can use a generalization of the AR (Anderson-Rubin) statistic introduced in Stock and Wright (2000):

$$AR(\alpha_0, \beta_0) = n \left(\frac{1}{n} \sum_i h(x_i) - M_n(\alpha_0, \beta_0) \right)' \Sigma^{-1} \left(\frac{1}{n} \sum_i h(x_i) - M_n(\alpha_0, \beta_0) \right),$$

where Σ is the covariance matrix of vector $h(x_i)$ (which we take to be nonsingular) or a consistent estimate thereof. Under the null, we have that $AR(\alpha_0, \beta_0) \Rightarrow \chi^2_k$. This result requires only that a central limit theorem hold for $h(x_i)$ and is fully robust towards

weak identification of β. An alternative test for the full parameter vector is suggested in Kleibergen (2005), which uses a score (LM) statistic.

To test a hypothesis with a strongly identified nuisance parameter, e.g. $H_0 : \beta = \beta_0$, we can use the AR statistic for a test on both α and β, minimized over the nuisance parameter α. In particular, we consider

$$AR(\beta_0) = \min_{\alpha} AR(\alpha, \beta_0).$$

Stock and Wright (2000) prove that under the null $AR(\beta_0) \Rightarrow \chi^2_{k-k_\alpha}$ regardless of the strength of identification of β. Interested readers may find a full proof of this result in Stock and Wright (2000): here, we instead show that the problem of testing $H_0 : \beta = \beta_0$ for a strongly identified nuisance parameter is asymptotically equivalent to that of testing a linear hypothesis in the Gaussian model described in the beginning of section 2.

To see that this is the case, define $\xi_n = \sqrt{n}\Sigma^{-1/2}(\frac{1}{n}\sum_i h(x_i) - M_n(\alpha_0, \beta_0))$. By the central limit theorem, $\xi_n \Rightarrow \xi \sim N(0, I_k)$. Let S_n be the image of the function

$$\begin{aligned} m_n(\alpha) =& \sqrt{n}\Sigma^{-1/2}(M_n(\alpha, \beta_0) - M_n(\alpha_0, \beta_0)) = \\ =& \sqrt{n}\Sigma^{-1/2}(\widetilde{M}(\alpha) - \widetilde{M}(\alpha_0)) + \Sigma^{-1/2}(M^*(\alpha, \beta_0) - M^*(\alpha_0, \beta_0)) = \\ =& \sqrt{n}\Sigma^{-1/2}(\widetilde{M}(\alpha) - \widetilde{M}(\alpha_0)) + O(||\alpha - \alpha_0||). \end{aligned} \tag{5}$$

The statistic $AR(\beta_0)$ is equal to $\rho^2(\xi_n, S_n)$. For any bounded set $\mathcal{B}$, the intersection $S_n \bigcap \mathcal{B}$ converges to the intersection of $\mathcal{B}$ with the k_α-dimensional linear sub-space S spanned by the columns of the Jacobian of $\widetilde{M}(\alpha)$ at point α_0. Indeed, according to equation (5) and the assumption that α is globally identified, one can easily show that the range of values of α such that $m_n(\alpha) \in S_n \bigcap \mathcal{B}$ is of order $1/\sqrt{n}$. Any regular manifold, however, is arbitrarily well approximated by its tangent space, which we can denote here by S, on an infinitesimal neighborhood of a regular point (see section 3.1 for definitions). As a result, it is easy to show that $\rho^2(\xi_n, S_n) \Rightarrow \rho^2(\xi, S) \sim \chi^2_{k-k_\alpha}$, where the last equality is true due to the fact discussed at the beginning of this section that the squared distance from a standard normal vector to a linear space passing through zero is χ^2-distributed. For another version of this asymptotic linearity result, see section 5, where we show that in models with strongly identified nuisance parameters the curvature is of order $1/\sqrt{n}$.

Tests for hypotheses with weakly identified nuisance parameters behave quite differently. In particular, we show that in general the curvature of a null hypothesis with a weakly identified nuisance parameter does not disappear asymptotically. To illustrate this point, assume that the hypothesis of interest is $H_0 : \alpha = \alpha_0$, so that β is a weakly identified nuisance parameter. Again, we consider the AR statistic minimized over the nuisance parameter:

$$AR(\alpha_0) = \min_{\beta} n \left(\frac{1}{n} \sum_i h(x_i) - M_n(\alpha_0, \beta) \right)' \Sigma^{-1} \left(\frac{1}{n} \sum_i h(x_i) - M_n(\alpha_0, \beta) \right).$$

Let us define $\xi_n = \sqrt{n}\Sigma^{-1/2}(\frac{1}{n}\sum_i h(x_i) - M_n(\alpha_0, \beta_0))$ as before, and let S_n be the image of the function

$$m_n(\beta) = \sqrt{n}\Sigma^{-1/2}(M_n(\alpha_0, \beta) - M_n(\alpha_0, \beta_0)) = \Sigma^{-1/2}(M^*(\alpha_0, \beta) - M^*(\alpha_0, \beta_0)).$$

By construction, S_n is a p-dimensional manifold in k-dimensional Euclidean space. In contrast to the strongly identified case, however, we have that S_n does not change with the sample size, so denote it S. Hence, if S_n is nonlinear for a given sample size, it remains nonlinear in the limit. As a result, we have that

$$AR(\alpha_0) = \rho^2(\xi_n, S) \Rightarrow \rho^2(\xi, S),$$

where $\xi \sim N(0, I_k)$ and S is a p-dimensional manifold, which is **not** in general a linear sub-space. Note, however, that this is precisely the problem discussed at the beginning of section 2. Hence, the problem of testing a hypothesis with weakly identified nuisance parameters using the AR statistic is asymptotically equivalent to that of testing a possibly nonlinear hypothesis in a Gaussian model. As a result, constructing bounds for the distribution of $\rho^2(\xi, S)$ will also allow us to conduct inference in models with weakly identified nuisance parameters.

Linearity vs strength of identification. We showed that the problem of testing a hypothesis with strongly identified nuisance parameters is asymptotically equivalent to that of testing that the mean belongs to linear subspace in a finite-dimensional Gaussian model. In contrast, if there are weakly identified nuisance parameters the manifold

corresponding to the null hypothesis need not converge to a linear subspace, so the usual critical values may be invalid. Rather than focusing on strength of identification, however, we may view the key distinction here as between linearity and non-linearity. In particular, while strong identification guarantees that the null hypothesis will correspond to a linear subspace in the limit, even with weakly identified nuisance parameters if S (the set of parameter values satisfying the null) happens to be a linear subspace, the usual $\chi^2_{k-k_\beta}$ limiting distribution will be correct. Hence, in models where the nuisance parameters enter the function M linearly, the usual (strong-identification) critical values for the AR statistic will yield asymptotically valid tests regardless of the strength of identification. Asymptotic linearity, not strong identification as such, is the essential condition. On a related note, Andrews and Mikusheva (2011) show that in a parametric model, a score test which concentrates out the nuisance parameter has asymptotically correct coverage even for testing a null hypothesis with a weakly identified nuisance parameter as long as this parameter enters the log-likelihood function linearly.

2.2 Minimal distance statistics

The analysis above generalizes to a broader minimum-distance context. Assume that we have a sample of size n from a model parameterized by structural parameters (α, β), which are related to reduced form parameters θ by $\theta = M_n(\alpha, \beta)$ where the function M_n is of the form described in equation (3). Further, suppose we have an estimator $\hat{\theta}$ of the reduced-form parameters which is consistent and asymptotically normal:

$$\sqrt{n}(\hat{\theta} - \theta) \Rightarrow N(0, \Sigma),$$

for Σ either known or consistently estimable: in the GMM example above, we could take $\theta = E[h(x_i)]$ and $\hat{\theta} = \frac{1}{n}\sum_{i=1}^n h(x_i)$. As before, α is strongly identified while β is weakly identified. The analysis now applies to so-called minimum distance statistics (which we will continue to refer to as AR for simplicity). In particular, to test the hypothesis $H_0 : \alpha = \alpha_0, \beta = \beta_0$ we use

$$AR(\alpha_0, \beta_0) = n\left(\hat{\theta} - M_n(\alpha_0, \beta_0)\right)' \Sigma^{-1} \left(\hat{\theta} - M_n(\alpha_0, \beta_0)\right),$$

while to test the hypothesis $H_0 : \beta = \beta_0$ with a strongly identified nuisance parameter we consider the statistic $AR(\beta_0) = \min_\alpha AR(\alpha, \beta_0)$ and pair it with $\chi^2_{k-k_\alpha}$ critical values. Taking $\xi_n = \sqrt{n}\Sigma^{-1/2}(\hat{\theta} - \theta_0)$, the same argument as above shows that this testing problem is asymptotically equivalent to testing a linear hypothesis in a Gaussian model. Likewise, testing a hypothesis with a weakly identified nuisance parameter is again asymptotically equivalent to testing a non-linear hypothesis in a Gaussian model.

Below we discuss several applied examples that can be cast into this setting.

2.2.1 Example: DSGE models

Dynamic Stochastic General Equilibrium (DSGE) models have recently been quite popular in applied Macroeconomics. These are highly non-linear, very multi-dimensional dynamic models describing the evolution of the main macro indicators in the economy and are used by many central banks. A number of concerns have been voiced about identification in these models (Ruge-Murcia (2007), Canova and Sala (2009), Iskrev (2011), I. Andrews and Mikusheva (2011), Guerron-Quintana, Inoue and Kilian (2009)), and many authors have noted that standard frequentist statistical procedures are unreliable. The source and extent of weak identification in such models is not well understood, and it is impossible to distinguish which parameters are weakly identified using currently-available procedures.

Several recent papers (Dufour, Khalaf and Kichian (2009), Guerron-Quintana, Inoue and Kilian (2009), I. Andrews and Mikusheva (2011), and Qu (2011)) suggest tests for full parameter vector hypotheses robust towards weak identification. With the exception of I. Andrews and Mikusheva (2011), these papers suggest the projection method for inference on subsets of parameters. Due to the high dimension of the parameter vector in many DSGE models, however, the projection method tends to be quite conservative.

Most DSGE models can be cast into our framework, which seems natural here, as one suggestion for how to estimate DSGE models is through two-step matching procedures (Christiano and Eichenbaum (1992), Rotemberg and Woodford (1997), Ruge-Murcia (2010)). Typical log-linearized DSGE models are of the form:

$$\Gamma_0(\beta) z_t = \Gamma_1(\beta) E_t z_{t+1} + \Gamma_2(\beta) z_{t-1} + \Gamma_3(\beta) u_t,$$

where z_t is a set of state variables at time t, β is a set of structural parameters, u_t are

i.i.d. mean zero shocks with identity covariance matrix, and $\Gamma_i(\beta)$ are known (often non-linear) functions. We observe variables $x_t = Cz_t$, where in many cases C is a deterministic matrix (usually a selection matrix). There are a number of known procedures for solving systems of linear rational expectation equations of this form- see Blanchard and Kahn (1980), Anderson and Moore (1985), King and Watson (1998), and Sims (2002) for different algorithms.

Once solved, the model can be written in the form

$$z_t = A(\beta)z_{t-1} + B(\beta)u_t,$$

where matrices $A(\beta)$ and $B(\beta)$ are generally nonlinear in β and solve the equations

$$(\Gamma_0 - \Gamma_1 A)A - \Gamma_2 = 0; \quad (\Gamma_0 - \Gamma_1 A)B - \Gamma_3 = 0.$$

In this context, a natural choice of reduced-form parameters is the auto-covariances of the observed vector-series x_t. In particular, let $\Sigma_x(j)$ be j-th order auto-covariance of x_t (for details see Iskrev (2010)):

$$\Sigma_x(j) = cov(x_t, x_{t-j}) = CA^j\Sigma_z(0)C'$$

where $\Sigma_z(0) = Ez_tz_t' = A\Sigma_z(0)A' + BB'$. It is helpful to write everything in vectorized form. In particular,

$$\theta_j(\beta) = vec(\Sigma_x(j)) = (C \otimes CA^j)(I - (A \otimes A))^{-1}vec(BB').$$

One may choose the reduced form parameter θ to be some subset of $vec(\Sigma_x(j))$, so $\theta = m(\beta) = W(\theta_0(\beta)', ..., \theta_j(\beta)')'$, where W is a selection matrix. In the absence of persistence (exact or near unit roots) the sample estimators

$$\hat{\theta}_j = vec\left(\frac{1}{T-j-1}\sum_{t=1}^{T-j}(x_{t+j} - \overline{x})(x_t - \overline{x})'\right)$$

of θ_j satisfy a central limit theorem and achieve normality quite quickly. As a result, normal approximations to the distribution of $\hat{\theta} = W(\hat{\theta}_0, ..., \hat{\theta}_j)$ are usually quite reliable for realistic sample sizes. Hence, we can conduct inference on the structural parameters

β using the AR statistic

$$AR(\beta) = n(\hat{\theta} - \theta(\beta))'\Sigma^{-1}(\hat{\theta} - \theta(\beta))$$

where Σ is the covariance matrix of the reduced-form parameter estimates.

2.2.2 Example: Phillips curve

Typical models of the New Keynesian Phillips curve relate present inflation to expectations about future inflation and past values of inflation. One such formulation is

$$\pi_t = \sum_{j=0}^{n} \lambda_j s_{t-j} + \gamma_f E_t \pi_{t+1} + \sum_{j=1}^{m} \gamma_j \pi_{t-j} + \varepsilon_t, \tag{6}$$

where π_t is the inflation at time t and s_t is some driving variable (e.g. labor costs). One way of estimating such models, popularized by Gali and Gertler (1999), is with instrumental variables regression.

Mavroeidis (2005) argues that models of forward-looking expectations (like the New Keynesian Phillips curve) require an analysis of identification distinct from the usual GMM-IV arguments considered in i.i.d. models. In particular, the error in the Phillips curve expression above will in general be autocorrelated and heterogeneous, raising identification issues above and beyond those faced in the i.i.d. case. Mavroeidis (2005) shows that the dynamics of the forcing variable s_t are extremely important for determining the identification of the model and that if these dynamics are insufficiently rich then the parameters in (6) will not be identified. He also argues that standard diagnostics for identification strength designed for i.i.d. models may be quite misleading when applied to Phillips curve estimation.

To cast New Keynesian Phillips curve estimation into our setting, suppose that the driving variable s_t is weakly exogenous and can be modeled as

$$s_t = \sum_{j=1}^{p} \rho_j s_{t-j} + \sum_{j=1}^{q} \phi_j \pi_{t-j} + v_t. \tag{7}$$

One can then solve the model described by (6) and (7) to obtain a reduced form solution

$$\pi_t = \sum_{j=0}^{l_s} \alpha_j s_{t-j} + \sum_{j=1}^{l_\pi} \delta_j \pi_{t-j} + \alpha_\varepsilon \varepsilon_t. \tag{8}$$

To test hypotheses on the structural parameters $\beta = (\lambda_0, ..., \lambda_n, \gamma_f, \gamma_1, ..., \gamma_m, \rho_1, ..., \rho_p, \phi_1, ..., \phi_q)$, we first estimate the reduced form parameters $\theta = (\alpha_0, ..., \alpha_{l_s}, \delta_1, ..., \delta_{l_\pi}, \rho_1, ..., \rho_p, \phi_1, ..., \phi_q)$ by OLS regressions (7) and (8). The function $\theta(\beta)$ connecting the structural parameters to the reduced form is given by the solution to the model and can be found in Mavroeidis (2005). Using this solution, we can test any hypothesis on the structural parameters by testing the implied hypothesis on the reduced form parameters. In particular, for a full parameter vector hypothesis $H_0 : \beta = \beta_0$ we consider the AR statistic:

$$AR(\beta_0) = n(\hat{\theta} - \theta(\beta_0))' \Sigma^{-1} (\hat{\theta} - \theta(\beta_0)),$$

where n is the sample size and Σ is a HAC-consistent estimator of the asymptotic variance of $\hat{\theta}$. Likewise, any other hypothesis about the structural parameters describes some manifold in the space of θ's and thus fits into the framework described in section 2.

3 Geometry

3.1 Manifolds, tangent spaces, curvature

In this paper we focus on regular manifolds embedded in k-dimensional Euclidean space with the usual Euclidean norm $\|\cdot\|$. A subset $S \subset \mathbb{R}^k$ is called a *p-dimensional regular manifold* if for each point $q \in S$ there exists a neighborhood V in $\mathbb{R}^k$ and a twice-continuously-differentiable map $\mathbf{x} : U \to V \bigcap S$ from an open set $U \subset \mathbb{R}^p$ onto $V \bigcap S \subset \mathbb{R}^k$ such that (i) $\mathbf{x}$ is a homeomorphism, which is to say it has a continuous inverse and (ii) the Jacobian $d\mathbf{x}_q$ has full rank. A mapping $\mathbf{x}$ which satisfies these conditions is called a parametrization or a system of local coordinates, while the set $V \bigcap S$ is called a coordinate neighborhood.

Note that the manifold S is defined as a set, rather than as a map. In keeping with this spirit, many of the statements below will be invariant to parametrization. Hence, if we have different parameterizations for the same manifold, which of them we use is

entirely a matter of convenience. In some problems it may be the case that there does not exist a global parametrization, that is a fixed mapping $\mathbf{x}$ satisfying the conditions above such that S is the image of $\mathbf{x}$. For instance, in Example 1 we defined S as the set of points q satisfying the $k-p$-dimensional restriction $g(q) = 0$ where the twice continuously differentiable function g had a Jacobian of rank $k-p$ at all points. Given this definition, the Implicit Function Theorem guarantees the existence of a local parametrization in a neighborhood of each point $q \in S$ but a global parametrization need not exist.

We begin by developing some geometrical concepts for the special case of a regular 1-dimensional manifold, also known as a curve. In particular, let S be a curve parameterized by $\alpha : (t_0, t_1) \to \mathbb{R}^k$ where α is twice continuously differentiable and (t_0, t_1) is an interval in $\mathbb{R}$. The *arc length* is defined as $s(t) = \int_{t_0}^{t} \|\dot{\alpha}(\tau)\| d\tau$. Without loss of generality, we can take α to be parameterized by arc length s, in which case at all points $\|\dot{\alpha}(s)\| = 1$ and the vector $\ddot{\alpha}(s)$ is perpendicular to $\dot{\alpha}(s)$. The vector $\dot{\alpha}(s)$ is called the tangent vector to S at $q = \alpha(s)$, while $\kappa(s) = \|\ddot{\alpha}(s)\|$ is called the *curvature* at q. The curvature measures how quickly the curve S deviates from its tangent line local to q, and the scaling is such that a circle of radius C has curvature $1/C$ at all points.

The change of variables from arbitrary parametrization t to the arc length s is not necessary for the calculation of curvature. In particular, as before let $\dot{\alpha}(t)$ and $\ddot{\alpha}(t)$ denote the first and second derivatives of α, now with respect to t. If we let $(\ddot{\alpha}(t))^{\perp}$ be the part of $\ddot{\alpha}(t)$ orthogonal to $\dot{\alpha}(t)$, then the curvature at $q = \alpha(t)$ is $\kappa(t) = \frac{\|(\ddot{\alpha}(t))^{\perp}\|}{\|\dot{\alpha}(t)\|^2}$. One can show that this definition of curvature is invariant to parametrization, and hence that in the special case of a curve parameterized by arc length it reduces to the definition given above.

These concepts can all be extended to general regular manifolds. Fixing a p-dimensional manifold S, for any curve $\alpha : (-\varepsilon, \varepsilon) \to S$ on S which passes through the point $q = \alpha(0) \in S$ the tangent vector $\dot{\alpha}(0)$ is called a tangent vector to S at q. For $\mathbf{x}$ a system of local coordinates at q, the set of all tangent vectors to S at q coincides with the linear space spanned by the gradient $d\mathbf{x}_q$ and is called the *tangent space* to S at q (denoted $T_q(S)$). While we have defined the *tangent space* using the local coordinates $\mathbf{x}$, as one would expect from its geometrical interpretation $T_q(S)$ is independent of the parametrization.

To calculate the curvature at q, consider a curve $\alpha : (t_0, t_1) \to S$ which lies in S and passes through $q = \alpha(0)$. Taking $T_q^{\perp}$ to be the $k-p$-dimensional linear space orthogonal

to $T_q(S)$ and $X = \dot{\alpha}(0) \in T_q(S)$ to be the tangent vector to α at q, define

$$\kappa_q(X, S) = \frac{\left\|(\ddot{\alpha}(0))^{\perp}\right\|}{\|X\|^2},$$

where $(W)^{\perp}$ stands for the projection of W onto the space $T_q^{\perp}$. One can show that $\kappa_q(X, S)$ depends on the curve α only through X. The measure of curvature we consider is

$$\kappa_q(S) = \sup_{X \in T_q(S)} \kappa_q(X, S) = \sup_{X \in T_q(S)} \frac{\left\|(\ddot{\alpha}(0))^{\perp}\right\|}{\|X\|^2}. \tag{9}$$

This measure of curvature is closely related to the Second Fundamental Tensor (we refer the interested reader to Kobayashi and Nomizu (1969, v.2, ch. 7)), and is equal to the maximal curvature over all geodesics passing through the point q. As with the curvature measure discussed for curves, (9) is invariant to the parametrization. Also analogous to the 1-dimensional case, if S is a p-dimensional sphere of radius C then for each $q \in S$ we have $\kappa_q(S) = 1/C$. Finally, if S is a linear subspace then its curvature is zero at all points.

How to calculate curvature in practice. Let S be a p-dimensional manifold in $\mathbb{R}^k$, and let $\mathbf{x}$ be a local parametrization at a point q, $q = \mathbf{x}(y^*)$. Denote the derivatives of $\mathbf{x}$ at q by $v_i = \frac{\partial \mathbf{x}}{\partial y_i}(y^*)$. By the definition of a local parametrization, we know that the Jacobian $Z = (v_1, ..., v_p)$ is full rank, so the tangent space $T_q(S) = span\{v_1, ..., v_p\}$ is p-dimensional. As before, for any vector $W \in \mathbb{R}^k$ let $W^{\perp}$ denote the part of W orthogonal to $T_q(S)$, that is, $W^{\perp} = (I - Z(Z'Z)^{-1}Z')W$. Finally, denote the p^2 vectors of second derivatives $V_{ij} = \frac{\partial^2}{\partial y_i \partial y_j}\mathbf{x}(y^*)$. The curvature can then be written as

$$\kappa_q(S) = \sup_{\substack{u=(u_1,..,u_p)\in\mathbb{R}^p \\ \|\sum_{i=1}^p u_i v_i\|=1}} \left\| \sum_{i,j=1}^{p} u_i u_j V_{ij}^{\perp} \right\| = \sup_{(w_1,...,w_p)\in\mathbb{R}^p} \frac{\left\|\sum_{i,j=1}^p w_i w_j V_{ij}^{\perp}\right\|}{\left\|\sum_{i=1}^p w_i v_i\right\|^2}. \tag{10}$$

3.2 Geometric bounds

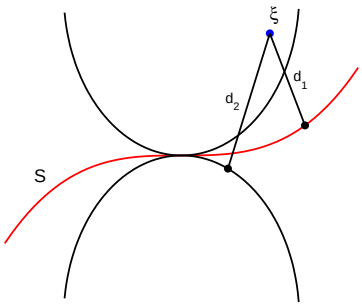
In this section we establish a bound on the distribution of the distance in $\mathbb{R}^k$ from a random point $\xi \sim N(0, I_k)$ to a p-dimensional non-random manifold S that contains

zero. Our bound depends on the maximal curvature $\kappa_q(S)$ over all relevant points in the manifold S. Our bound will be based on global properties of the manifold, in the sense of properties that hold on a fixed bounded set, but we abstract from the behavior of the manifold at infinity as irrelevant. In what follows, we restrict attention to a connected part of the manifold that lies inside of a (large) finite cylinder centered at zero.

We derive our bound in two steps: first, we construct an envelope for the manifold S using a collection of p-dimensional spheres. We show that the distance from any point ξ to S is bounded above by the distance from ξ to the most distant sphere in the collection we consider. Second, we show that our geometric construction implies a bound on the distribution of $\rho^2(\xi, S)$ and hence on the distribution of AR (or LR) statistics. To provide intuition for our main statement we walk the reader through two simple cases in which the construction of the envelope can be easily visualized.

Case 1 (k=2, p=1): A curve in $\mathbb{R}^2$. Consider a curve S passing through zero (i.e. $(0,0) \in S$). Suppose that the curvature of S is less than or equal to $1/C$ for all points in S. If we imagine two circles of radius C tangent to S at zero, we can see that the curve lies between them- see Figure 1 for illustration. Since S lies between the circles, the distance from any point ξ to S (denoted by d_1 in Figure 1) does not exceed the distance from ξ to the further of the two circles (denoted by d_2). This is the geometrical bound we use. Note that if the maximal curvature of S goes to zero at all points (so that $C \to \infty$)

Figure 1: Bounding a line between two circles.



then the two bounding circles converge to the tangent line to S at zero on any bounded set. Further, note that the distribution of the distance d_2 from a normal random vector to the furthest of two circles depends only on C and is easy to simulate.

The logic of this example is quite straightforward to generalize to the case of a $k-1$-

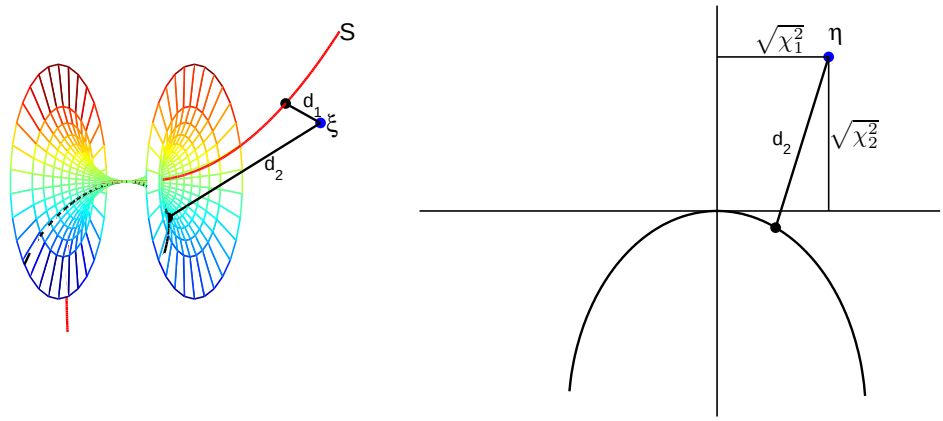


Figure 2: Left panel: the envelope for a space curve in $\mathbb{R}^3$. Right panel: distribution of the distance d_2

dimensional manifold in $\mathbb{R}^k$, known as a hyper-surface or a manifold of co-dimension 1.[4] If a regular $k-1$ dimensional manifold S in $\mathbb{R}^k$ has curvature $\kappa_q(S) \leq 1/C$ at all $q \in S$, consider the two $k-1$-dimensional spheres of radius C which are tangent to the manifold at zero. One can show (see Theorem 1 below) that as in the one-dimensional case S lies between these two spheres. Hence, we again have that the distance from any point ξ to S is bounded above by the distance from ξ to the furthest of the two spheres. Likewise, if the maximal curvature of S goes to zero (so that $C \to \infty$) we again have that on any bounded set the two spheres converge to the tangent space to S at zero, which in this case is a $k-1$-dimensional hyperplane.

Dealing with manifolds of co-dimension greater than 1 is much more challenging, but the basic principle of the approach can be illustrated using a curve in $\mathbb{R}^3$.

Case 2 (k=3, p=1): A curve in $\mathbb{R}^3$. Suppose now that we have a one-dimensional space curve S in $\mathbb{R}^3$ which passes through zero and whose curvature at all points is bounded above by $1/C$. We construct our envelope by considering the collection of all one-dimensional circles of radius C tangent to S at zero. Equivalently, one can take a given circle tangent to S at zero and rotate it around the tangent line. An example of the resulting surface is given on the left panel of Figure 2: as in the case of co-dimension 1, we can see that the curve S lies inside the envelope. One can show that the distance from any point ξ to the curve S (denoted by d_1 in Figure 2) is bounded above by the distance from ξ to the furthest circle in the collection used to construct the envelope

[4]The co-dimension of a manifold is the difference between the dimension of the space and the dimension of the manifold.

(denoted d_2). Note that if the curvature of S goes to zero at all points (so that $C \to \infty$) then on any bounded set the envelope we consider converges to the tangent line to S at zero.

This geometric bound immediately implies a bound on the distribution of $\rho(\xi, S)$. For $\xi \sim N(0, I_3)$ the distribution of the distance d_2 from ξ to the furthest circle is quite simple to simulate. One can show that it is distributed as the distance from two-dimensional random vector η depicted on the right panel of Figure 2 to the circle of radius C with center $(0, -C)$ where the coordinates of η are distributed as independent $\sqrt{\chi_1^2}$ and $\sqrt{\chi_2^2}$ random variables.

General case With the intuition provided by these examples, we now turn to the general case. Let S be a regular connected p-dimensional manifold in $\mathbb{R}^k$ passing through zero. By the rotation invariance of standard normal vectors we can assume without loss of generality that the tangent space $T_0(S)$ to manifold S at zero is spanned by first p basis vectors. For each $x \in \mathbb{R}^k$, let $x = (x^{(1)}, x^{(2)})$ where $x^{(1)} = (x_1, ..., x_p) \in \mathbb{R}^p$ contains the first p coordinates of x while $x^{(2)} = (x_{p+1}, ..., x_k) \in \mathbb{R}^{k-p}$ contains the last $k - p$. In what follows, we restrict attention to points on the manifold that lie inside of a (large) finite cylinder $D_C = \{x = (x^{(1)}, x^{(2)}) : \|x^{(1)}\| \leq C, \|x^{(2)}\| \leq C, x^{(1)} \in \mathbb{R}^p, x^{(2)} \in \mathbb{R}^{k-p}\} \subset \mathbb{R}^k$. Let S_C be the intersection $S \bigcap D_C$ if it is connected or the *connected* part of $S \bigcap D_C$ that passes through zero (that is, the part of $S \bigcap D_C$ which can be reached by continuous paths lying in $S \bigcap D_C$ which pass through zero) if $S \bigcap D_C$ is not connected. Note that $\rho(\xi, S) \leq \rho(\xi, S_C)$.

To obtain some of our bounding results, we need one further assumption:

Assumption 1 *For any $y^{(1)} \in \mathbb{R}^p$ with $\left\|y^{(1)}\right\| \leq C$ there exists a point $x \in S_C$ such that $x^{(1)} = y^{(1)}$.*

Assumption 1 requires that the projection of S on the tangent space to S at zero covers some fixed p-dimensional ball, and hence that S has dimension p in a global sense. By a local property we mean one that holds on an infinitesimal neighborhood of a point. In contrast, by a global property we mean one that holds on a fixed bounded set. We have already imposed a local dimensionality assumption on S by restricting the rank of the tangent space at all points. The distribution of the AR statistic, however, depends on global properties of the manifold S and so to bound the distribution we need a

global dimensionality assumption. To illustrate why local dimensionality assumptions are insufficient, imagine a strip $S = \{(x, y, z) \in \mathbb{R}^3 : z = 0, -\varepsilon < y < \varepsilon, x \in \mathbb{R}\}$ in $\mathbb{R}^3$. At any point $q \in S$ the dimension of the tangent space is equal to 2, but if $\varepsilon > 0$ is small enough then S does not satisfy Assumption 1. For ε sufficiently small, however, the distance from ξ to S behaves like the distance from ξ to the line $S^* = \{(x, y, z) : y = 0, z = 0\}$, which is one dimensional both locally and globally.

Theorem 1 *Let S be a regular p-dimensional manifold in $\mathbb{R}^k$ passing through zero. Assume that the tangent space $T_0(S)$ is spanned by first p basis vectors. Assume that for some constant $C > 0$ we have that $\kappa_q(S) < \frac{1}{C}$ for all points $q \in S_C$. Then:*

(a) Manifold S_C lies inside the set $\mathcal{M} \cap D_C$, where

$$\mathcal{M} = \{\|x^{(1)}\|^2 + (C - \|x^{(2)}\|)^2 \geq C^2\}. \tag{11}$$

(b) If Assumption 1 is satisfied, then for any point $\xi \in \mathbb{R}^k$ we have

$$\rho(\xi, S) \leq \max_{u \in \mathbb{R}^{p-k}, \|u\|=1} \rho(\xi, N_u),$$

where $N_u = \{x \in \mathbb{R}^k : x = (x^{(1)}, zu), x^{(1)} \in \mathbb{R}^p, z \in \mathbb{R}_+, \|x^{(1)}\|^2 + (C - z)^2 = C^2\}$.

(c) $\max_{u \in \mathbb{R}^{p-k}, \|u\|=1} \rho(\xi, N_u) = \rho(\xi, N_{\widetilde{u}})$, where $\widetilde{u} = -\frac{1}{\|\xi^{(2)}\|}\xi^{(2)}$.

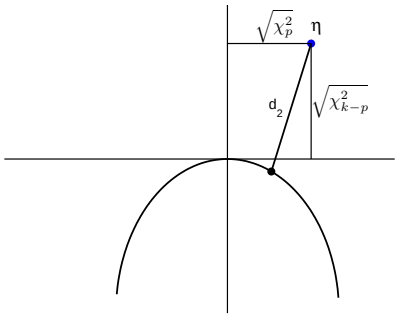
(d) If $\xi \sim N(0, I_k)$ we have for all x, y:

$$P\left\{\max_{u \in \mathbb{R}^{p-k}, \|u\|=1} \rho^2(\xi, N_u) \leq x, \|\xi\| \leq y\right\} = P\left\{\rho_2^2(\eta, N_2^C) \leq x, \|\eta\| \leq y\right\},$$

where the coordinates of the 2-dimensional random vector $\eta = (\sqrt{\chi_p^2}, \sqrt{\chi_{k-p}^2}) \in \mathbb{R}^2$ are independently distributed, $N_2^C = \{(z_1, z_2) \in \mathbb{R}^2 : z_1^2 + (C + z_2)^2 = C^2\}$ is a circle of radius C with the center at $(0, -C)$, and ρ_2 is Euclidian distance in $\mathbb{R}^2$.

Theorem 1 (a) establishes that the manifold S_C lies inside the set $\mathcal{M}$ bounded by an envelope we construct from a collection of p-dimensional spheres N_u. Statement (b) asserts that the distance from a point ξ to the manifold S is bounded by the distance from ξ to the furthest sphere in this collection, while (c) picks out exactly which sphere $N_{\widetilde{u}(\xi)}$ is the furthest away for a given ξ. Finally, (d) shows that the distribution of the

Figure 3: The stochastic bound described in Theorem 1 (d).



distance from $\xi \sim N(0, I_k)$ to $N_{\tilde{u}(\xi)}$ is the same as the distribution of the distance from a random variable η to a particular circle in $\mathbb{R}^2$ as depicted in Figure 3.

3.3 Stochastic bound

Theorem 1 implies a bound on the distribution of the distance from $\xi \sim N(0, I_k)$ to a p-dimensional manifold S. Assume that for some $C > 0$, S satisfies all the assumptions of Theorem 1 including Assumption 1. Then almost surely,

$$\rho^2(\xi, S) \leq \rho^2(\xi, N_{\tilde{u}}), \tag{12}$$

as follows from statement (b) of Theorem 1. By Theorem 1 (d), the distribution of the right hand side of (12) is the same as the distribution of the random variable variable ψ_C defined as

$$\psi_C = \rho_2^2(\eta, N_2^C), \tag{13}$$

where the coordinates of the two-dimensional random vector $\eta = (\sqrt{\chi^2_p}, \sqrt{\chi^2_{k-p}}) \in \mathbb{R}^2$ are independently distributed, $N_2^C = \{(z_1, z_2) \in \mathbb{R}^2 : z_1^2 + (C + z_2)^2 = C^2\}$ is a circle of radius C with the center at $(0, -C)$, and ρ_2 is Euclidean distance in $\mathbb{R}^2$. Combining these results, we establish the bound

$$P\left\{\rho^2(\xi, S) \geq x\right\} \leq P\left\{\psi_C \geq x\right\} \qquad \text{for all } x > 0,$$

so the distribution of ψ_C is an upper bound on the distribution of $\rho^2(\xi, S)$. We make the following observations:

(1) The distribution of ψ_C depends only on the dimension of the space k, the dimension of the manifold p and the maximal value of the curvature, $\frac{1}{C}$.

(2) The distribution of ψ_C is stochastically increasing in the maximal curvature and hence stochastically decreasing in C, so if $C_1 < C_2$ then ψ_{C_1} first-order stochastically dominates ψ_{C_2}.

(3) $\psi_C \Rightarrow \chi^2_{k-p}$ as $C \to \infty$, so if the curvature converges to zero at all relevant points then our bounding distribution converges to distribution of the distance from $\xi \sim N(0, I_k)$ to a p-dimensional linear subspace.

(4) At the other extreme, $\psi_C \Rightarrow \chi^2_k$ as $C \to 0$ so if the curvature of the manifold becomes arbitrarily large our bound coincides with the naive bound (2) that can be imposed without any assumptions on the manifold.

We want to emphasize that what we suggest is a stochastic bound that holds under quite general assumptions. If the model of interest has additional structure, this can potentially be exploited to obtain tighter bounds.

3.4 Statistical application of the stochastic bound

Suppose we have a single observation $\widehat{\theta}$ from a population $\widehat{\theta} \sim N(\theta_0, \Sigma)$ with an unknown mean θ_0. We wish to test a hypothesis of the form $H_0 : \theta_0 = \theta(\beta)$ for some value of the p-dimensional structural parameter $\beta \in U \subset \mathbb{R}^p$. As discussed above, this problem may arise when we have asymptotically normal reduced form estimates and are interested in testing hypotheses on the structural parameters as in sections 2.2.1 and 2.2.2. We use the AR statistic

$$AR = \min_{\beta}(\widehat{\theta} - \theta(\beta))'\Sigma^{-1}(\widehat{\theta} - \theta(\beta)) = \rho^2(\xi, S), \tag{14}$$

where $\xi = \Sigma^{-1/2}(\widehat{\theta} - \theta_0) \sim N(0, I_k)$ is an appropriately normalized version of our reduced form estimate $\hat{\theta}$ and the manifold $S = \{\Sigma^{-1/2}(\theta(\beta) - \theta_0), \beta \in U \subset \mathbb{R}^p\} \subset \mathbb{R}^k$ describes the restrictions imposed on the reduced-form parameters by the tested hypothesis and passes through zero if the null is true. If the manifold S satisfies the assumptions of Theorem 1 then by the argument in Section 3.3 the AR statistic is stochastically dominated by ψ_C

under the null, so if we use the $(1-\alpha)-$quantile of the distribution of ψ_C (which is easy to simulate) as a critical value the resulting test has size not exceeding α.

A practical question is what value of C to use. According to Theorem 1, the value of C is tied to the maximum of the curvature of S over the intersection of S with a cylinder D_C centered at zero. Notice, however, that in practice we do not observe the manifold S, since it depends on the unknown θ_0. However, the desired curvature is the same as the maximal curvature of the manifold $\widetilde{S} = \{\Sigma^{-1/2}\theta(\beta), \beta \in U \subset \mathbb{R}^p\} \subset \mathbb{R}^k$ over all points in the intersection of $\widetilde{S}$ with the cylinder $\widetilde{D}_C(x_0) = \{x \in \mathbb{R}^k : x - x_0 \in D_C\}$ centered at $x_0 = \Sigma^{-1/2}\theta_0$. This maximal curvature, in turn, is clearly bounded above by the maximal curvature over the whole manifold, so if we take $\widetilde{C} = 1/\left(\max_{q\in\widetilde{S}} \kappa_q(\widetilde{S})\right)$, using critical values based on $\psi_{\widetilde{C}}$ provides a test that controls size. Moreover, since $\widetilde{C}$ does not depend on any unobservables, a test based on these critical values is feasible.

If the null hypothesis has a global parametrization, as when $H_0 : \theta_0 = \theta(\beta), \beta \in U$, let $\kappa(\beta) = \kappa_{q=\theta(\beta)}(S)$. The latter is a function on U which depends only on the first two derivatives of $\theta(\beta)$. Hence, if we can evaluate these derivatives finding $\widetilde{C} = \max_\beta \kappa(\beta)$ is a standard non-stochastic optimization problem. If $\theta(\beta)$ is fairly tractable we may be able to solve for $\widetilde{C}$ analytically, while if not we can use the usual menu of numerical optimization techniques, such as Newton's method.

There are a variety of problems, however, in which using $\widetilde{C}$ may be unappealing. For example, it may be that calculating derivatives of $\theta(\beta)$ is challenging, or that the manifold has irregularities or points of high curvature which are far away from $\widehat{\theta}$. In such cases we may wish to restrict attention to the curvature of the manifold over some smaller set, which raises two issues. First, we do not know the true value θ_0 and hence the center of the cylinder $\widetilde{D}_C(x_0)$. Second, if the manifold is close to flat (so C is large) to find the maximal curvature over $\widetilde{D}_C(x_0)$ we might need to check the curvature over a huge set, which could be very computationally demanding.

We suggest a test which overcomes both of these problems and is easy to implement in practice. For a fixed value R, let $C \wedge R = \min\{C, R\}$. Denote by $F_\alpha(C, R, k, p)$ the $\alpha-$quantile of the distribution of $\psi_C(R)$ defined as

$$\psi_C(R) = \begin{cases} \rho_2^2(\eta, N_2^C) & \text{if } \|\eta\| \leq R; \\ \|\eta\|^2 & \text{if } \|\eta\| > R, \end{cases}$$

where η and and N_2^C are defined in statement (d) of Theorem 1. For any finite R the distribution of $\psi_C(R)$ provides a weaker bound than the distribution of ψ_C. This is the price paid for calculating curvature over a smaller set of points.

Lemma 1 *Assume that we have a single observation $\widehat{\theta}$ from a population $\widehat{\theta} \sim N(\theta_0, \Sigma)$ with unknown mean θ_0. We wish to test the hypothesis $H_0 : \theta_0 = \theta(\beta)$ for some $\beta \in U \subset \mathbb{R}^p$. Let $\widetilde{S} = \{\Sigma^{-1/2}\theta(\beta), \beta \in U \subset \mathbb{R}^p\} \subset \mathbb{R}^k$ be a regular p-dimensional manifold, and $\widetilde{B} = B_{(1+\sqrt{2})R}(\widehat{x})$ a ball of radius $(1+\sqrt{2})R$ around $\widehat{x} = \Sigma^{-1/2}\widehat{\theta}$, where R is such that $P\{\chi_k^2 \geq R^2\} < \alpha$. Let*

$$\widehat{C} = \begin{cases} \left(\min_{q \in \widetilde{S} \bigcap \widetilde{B}} 1/\kappa_q(\widetilde{S})\right) \wedge R, & \text{if } \widetilde{S} \bigcap \widetilde{B} \neq \emptyset; \\ 0, & \text{if } \widetilde{S} \bigcap \widetilde{B} = \emptyset. \end{cases}$$

Assume that for any $x \in \widetilde{S}$ such that $\|x - \widehat{x}\| \leq R$ we have that the projection of $\widetilde{S} \bigcap B_R(x)$ onto $T_x(\widetilde{S})$ contains a p-dimensional ball centered at x with radius $\widehat{C} \wedge R$. Then the test which rejects the null if and only if $AR = \min_\beta (\widehat{\theta} - \theta(\beta))' \Sigma^{-1} (\widehat{\theta} - \theta(\beta)) > F_{1-\alpha}(\widehat{C}, R, k, p)$ has size not larger than α.

4 Comparison with other methods available for testing hypotheses with weak nuisance parameters

As previously discussed, there is a wide literature devoted to the problem of weak-identification-robust tests for the full parameter vector and for hypotheses with strongly identified nuisance parameters, but much less is known about testing with weakly identified nuisance parameters.

Projection method. Recently, the projection method has been the standard approach to inference with weakly identified nuisance parameters. The projection method was introduced and popularized in econometrics by Dufour and Jasiak (2001) and Dufour and Taamouti (2005), and recent applications to non-standard testing problems in econometrics include Dufour, Khalaf, and Kichian (2006), Guerron-Quintana, Inoue and Kilian (2009), and Qu (2011).

The projection method is based on the observation that the hypothesis $H_0 : \alpha = \alpha_0$

with nuisance parameter β is equivalent to the hypothesis

$$H_0 : \exists \ \ \beta_0 \text{ s.t. } \alpha = \alpha_0, \beta = \beta_0.$$

Hence, to test a hypothesis on α we can use the statistic $AR(\alpha_0) = \inf_{\beta^*} AR(\alpha_0, \beta^*)$, and since

$$AR(\alpha_0) = \inf_{\beta^*} AR(\alpha_0, \beta^*) \leq AR(\alpha_0, \beta_0) \sim \chi^2_k,$$

we know that using χ^2_k critical values will yield a test which controls size. The name "projection method" stems from the fact that constructing confidence sets for α with this procedure is equivalent to constructing a joint confidence set for (α_0, β_0) using the full-vector AR statistic and then projecting this set on the parameter space for α.

The obvious advantage of the projection method is that it requires no assumptions about the strength of identification of β, since it relies only on the validity of the test for the full parameter vector. Other advantages include that it is quite easy to use and very broadly applicable. The primary disadvantage of the projection method is its conservativeness. Our test, introduced in Section 3.4, is based on the same statistic as the projection method (AR minimized over the nuisance parameters) but uses smaller critical values while still maintaining size. Only in the limiting case of infinitely high curvature ($C = 0$) do our critical values correspond to those of the projection method. As a result, except for this limiting case our test is strictly more powerful than the projection method and produces strictly smaller confidence sets in all realizations of the sample. Further, all of the assumptions we impose on the manifold S can be directly verified using the non-stochastic manifold $\widetilde{S}$ known to the researcher.

Concentrating out nuisance parameters. If one knows that the nuisance parameter β in a given testing problem is strongly identified then he/she can simply "concentrate out" the nuisance parameter, minimizing the AR statistic over β and reducing the degrees of freedom for the limiting distribution by k_β (that is, using quantiles of a $\chi^2_{k-k_\beta}$ rather than a χ^2_k). As discussed in section 2.1, this reduction in degrees of freedom stems from the fact that any manifold corresponding to a hypothesis with a strongly identified nuisance parameter converges to a linear subspace asymptotically.

The obvious advantage of this approach is that it is strictly more powerful than the projection method. However, the assumption of strong identification of the nuisance

parameter is essential, and the test may over-reject if this assumption fails. In many practical settings, including the DSGE and Phillips curve examples discussed above, the exact nature and source of weak identification is not clear, and we are unaware of any test of the null of weak identification which can be used to separate the weakly and strongly identified parameters. In contrast, the test we suggest in Section 3.4 does not employ any assumptions about the strength of identification of any parameter in point identified models. Indeed, since our approach is based on a finite-sample perspective (except perhaps for the assumed normality of the reduced-form parameters) we do not even require that there be a meaningful distinction between the weakly and strongly identified structural parameters in the model.

4.1 Pre-test for weak identification

If for some reason a researcher does not want to use our test, we suggest a simple procedure which could be called a "pre-test for weak identification." Imagine that a researcher wants to use a robust procedure (for example our test or the projection method) unless she knows that identification issues will not cause large size distortions, in which case she prefers instead to concentrate out the nuisance parameters. Our stochastic bounds can be used to address this question and determine whether weak identification constitutes a problem in a given setting. Below, we suggest a procedure which, when used as the first step of a two-step testing procedure of this sort, ensures that the procedure as a whole controls size.

To proceed, let us introduce the notion of a "tolerance level". Suppose that we would like to have a test of size α, but we are uncertain whether the usual strong-identification asymptotics provide a reasonable approximation in our context; in the event that these approximations are imperfect, we are willing to accept a test with true size $\alpha + \alpha^*$ in exchange for the additional power and convenience of using conventional critical values. The potential increase in the size α^* is called the tolerance level and has been used previously by e.g. Stock and Yogo (2005). For t-tests in weak IV, Stock and Yogo (2005) suggest comparing the minimal eigenvalue of the first-stage F statistic matrix to an appropriate threshold: if the minimal eigenvalue exceeds the threshold then the researcher can be confident that the usual 5% t-tests will have true size not exceeding 10%.

The pretest we propose asks whether the curvature of the model is sufficiently small to ensure that tests based on classical $\chi^2_{k-k_\beta}$ critical values (that is, tests which concentrate out the nuisance parameter β) with nominal size α have true size not exceeding $\alpha + \alpha^*$. To determine whether this is the case we calculate C^*, the smallest value C such that the $(1-\alpha)$-quantile of a χ^2_{k-p} distribution does not exceed $F_{1-(\alpha+\alpha^*)}(C, R, k, p)$. The cut-off C^* depends on the dimension k of the reduced-form parameter vector, the dimension p of the nuisance parameter, and R. To implement the pre-test we then calculate the value $\widehat{C}$ as described in Lemma 1, that is, the maximal curvature of the manifold at points inside a ball of radius $(1+\sqrt{2})R$ around the reduced-form estimator (noting that we may take $R = \infty$) and compare $\widehat{C}$ to C^*. If $\widehat{C} > C^*$ the researcher can safely concentrate out β and use $\chi^2_{k-k_\beta}$ critical values while if $\widehat{C} \leq C^*$ she should use a robust procedure. We can guarantee that the resulting two-step test will have size less than $\alpha + \alpha^*$.

Table 1 reports the cut-offs C^* for nominal 5% tests and tolerance level 5% for different values of p and k for R equal to the 0.99 quantile of a $\sqrt{\chi^2_k}$. Based on Table 1 we can see that for a fixed dimension k of the reduced-form parameter, increasing the number of nuisance parameters p tightens the restrictions imposed on curvature if one wants to concentrate out the nuisance parameters.

The main difference of this pre-test procedure from majority of existing tests of weak identification is that the resulting two-step procedure controls size. Tests of weak identification by Inoue and Rossi (2011), Iskrev (2010) and Wright (2003) all test the null of strong identification against the alternative of weak identification. Those tests control the probability of falsely rejecting strong identification but do not control the probability of failing to detect weak identification when it is present (this depends on power for these tests). As a result, using one of these tests in a two-step procedure, in which one tests the null of strong identification and uses a robust procedure only if strong identification is rejected, does not guarantee overall size control. In contrast, the procedure suggested here controls the size under weak identification.

4.2 AR conjecture

Kleibergen and Mavroeidis (2009a) consider a weak IV model with more than one endogenous regressor in which one wants to test a hypothesis about the coefficient on one endogenous regressor, treating the coefficients on the remaining regressors as nuisance

p\k	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.73	0.28	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2		2.33	1.94	1.65	1.41	1.20	1.00	0.82	0.65	0.49	0.34	0.19	0.05	0.00	0.00	0.00	0.00	0.00	0.00
3			4.00	3.63	3.36	3.13	2.93	2.74	2.57	2.41	2.26	2.11	1.97	1.84	1.71	1.59	1.46	1.35	1.23
4				5.72	5.38	5.11	4.88	4.68	4.50	4.34	4.18	4.03	3.89	3.76	3.63	3.50	3.38	3.27	3.15
5					7.46	7.16	6.90	6.68	6.47	6.29	6.13	5.97	5.83	5.69	5.55	5.43	5.31	5.19	5.07
6						9.23	8.96	8.71	8.49	8.29	8.10	7.94	7.78	7.63	7.49	7.36	7.24	7.11	7.00
7							11.00	10.78	10.55	10.32	10.12	9.94	9.77	9.61	9.46	9.32	9.19	9.06	8.94
8								12.80	12.62	12.40	12.18	11.98	11.79	11.62	11.46	11.31	11.17	11.03	10.90
9									14.58	14.46	14.25	14.05	13.85	13.66	13.48	13.32	13.17	13.03	12.89
10										16.38	16.31	16.12	15.92	15.72	15.54	15.37	15.20	15.05	14.90
11											18.18	18.16	18.00	17.80	17.61	17.43	17.25	17.08	16.93
12												19.97	20.02	19.88	19.69	19.51	19.32	19.15	18.98
13													21.78	21.88	21.76	21.59	21.41	21.22	21.05
14														23.58	23.75	23.65	23.49	23.31	23.14
15															25.38	25.62	25.55	25.40	25.23
16																27.19	27.49	27.45	27.31
17																	29.01	29.37	29.36
18																		30.83	31.25
19																			32.65

Table 1: Cut-off levels of C^* described in Section 4.1 for different values of k and p. Value of R is equal to 0.99- quantile of distribution χ^2_k. Based on 10 million simulations

parameters. They consider the AR statistic minimized over the nuisance parameters and show that its distribution under weak identification is dominated by its distribution under strong identification, and thus that χ^2 critical values with reduced degrees of freedom will always produce tests which maintain size. They also conjectured[5] that this statement could be generalized to *some* non-linear GMM models. If this statement held in all GMM models it would have eliminated the trade-off between projecting over and concentrating out nuisance parameters and implied that one should concentrate out in all cases to obtain more powerful tests while still controlling size. Below we provide a simple example which demonstrates that such a dominance result does not hold generally.

To fix ideas, assume that we have an i.i.d. sample $x_1, ..., x_n$ from a population that satisfies the k-dimensional moment condition $Ef_n(x_i, \theta) = 0$ at $\theta = \theta_0$ and assume that $\Sigma = \text{var}(f_n(x_i, \theta))$ is known. We first consider the AR statistic for testing the full parameter hypothesis $H_0 : \theta = \theta_0$:

$$AR(\theta_0) = \frac{1}{n}\left(\sum_{i=1}^{n} f_n(x_i, \theta_0)\right)\Sigma^{-1}\left(\sum_{i=1}^{n} f_n(x_i, \theta_0)\right)$$

This statistic is used with χ^2_k critical values. Assume that $\theta = (\alpha', \beta')'$ and we want to test the hypothesis $H_0 : \alpha = \alpha_0$ with nuisance parameter β. Consider the statistic:

$$AR(\alpha_0) = \min_{\beta} AR(\alpha_0, \beta).$$

We show that it is not in general true that the asymptotic distribution of $AR(\alpha_0)$ is stochastically dominated by a $\chi^2_{k-k_\beta}$, that is the limit distribution of $AR(\alpha_0)$ if β is strongly identified.

A counterexample to the AR conjecture. Consider an i.i.d. sample $x_1, ..., x_n$ drawn from $N(M_n(\theta_0), I_k)$ with k-dimensional structural parameter $\theta \in \Theta = [0, \pi]^{k-2} \times$

[5] The statement was made in one of the earlier versions of Kleibergen and Mavroeidis (2009a) and via private communication.

$[0, 2\pi) \times \mathbb{R} \subset \mathbb{R}^k$. Suppose further that

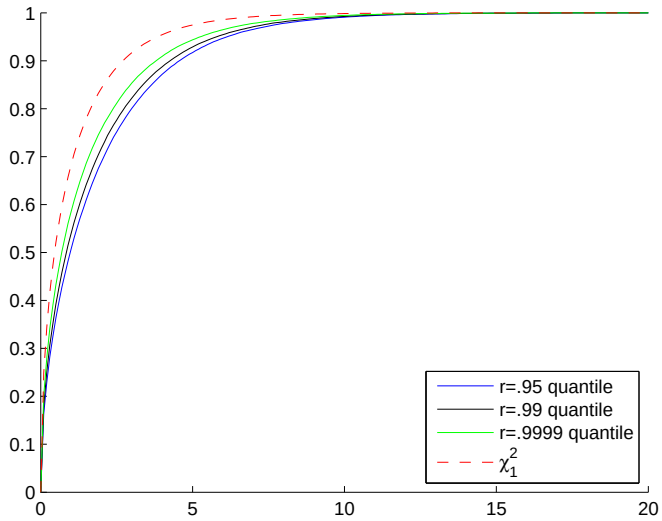
$$M_n(\theta) = \frac{r}{\sqrt{n}}\theta_k \begin{bmatrix} \cos(\theta_1) - 1 \\ \sin(\theta_1)\cos(\theta_2) \\ \sin(\theta_1)\sin(\theta_2)\cos(\theta_3) \\ \vdots \\ \sin(\theta_1)...\sin(\theta_{k-2})\cos(\theta_{k-1}) \\ \sin(\theta_1)...\sin(\theta_{k-2})\sin(\theta_{k-1}) \end{bmatrix}$$

is a k-dimensional vector-function of k variables. Note that all parameters are weakly identified in this case. The AR statistic for the full parameter vector is then

$$AR(\theta) = \frac{1}{n}\left(\sum_{i=1}^{n}\left[x_i - M_n(\theta)\right]\right)' \left(\sum_{i=1}^{n}\left[x_i - M_n(\theta)\right]\right).$$

Now assume that the hypothesis of interest is $H_0 : \theta_k = 1$, where $\beta = (\theta_1, ..., \theta_{k-1})'$ is a weakly identified nuisance parameter. The AR conjecture suggests that the limiting distribution of $\min_\beta AR(\beta, 1)$ is stochastically dominated by a χ^2_1. As before let us intro-

Figure 4: Distribution of statistic $\min_\beta AR(\beta, 1)$ for $k = 10$



duce a random variable $\xi = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[x_i - M_n(\theta_0)\right] \sim N(0, I_k)$, whose distribution does not depend on the sample size, and a function

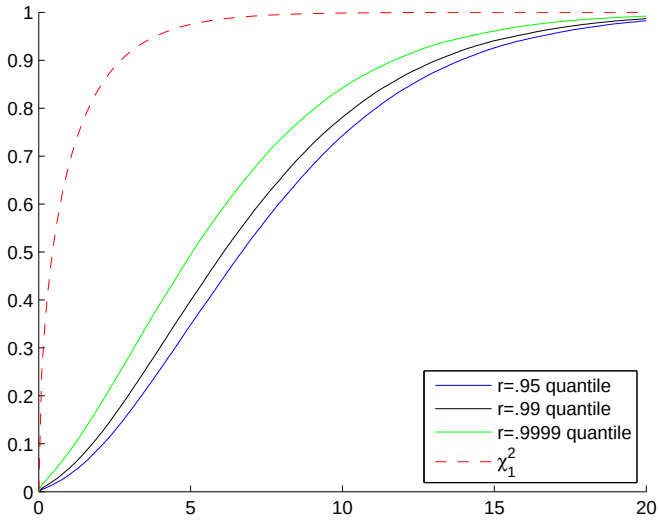
$$m(\beta) = \sqrt{n}\left(M_n(\beta, 1) - M_n(\theta_0)\right),$$

which also does not depend on the sample size. Notice that for all n the manifold described by the function $m(\beta)$ is a hyper-sphere of radius $r/\sqrt{n}$ in k-dimensional space passing through the origin. It is easy to see that

$$\min_{\beta} AR(\beta, 1) = \min_{\beta} \left(\xi - m(\beta)\right)' \left(\xi - m(\beta)\right).$$

Note, however, that the behavior of the last statistic does not depend on the sample size and can be easily simulated. Note further that since the distribution of a standard normal vector is rotation invariant, the distribution of $\min_\beta AR(\beta, 1)$ under the null depends only on k and the radius r. We simulate the distribution of the statistic of interest under the null for $k = 10$ and $k = 50$ and r equal to the square roots of the .95, .99, and .9999 quantiles of a χ^2_k distribution, and plot the resulting cdfs against a χ^2_1 cdf. Figures 4 and 5 show that tests which pair the minimized AR statistic with χ^2_1 critical values overreject. Further, the degree of over-rejection is increasing with the dimension k.

Figure 5: Distribution of statistic $\min_\beta AR(\beta, 1)$ for $k = 50$



4.3 Other methods

There are very few other papers that work directly with weakly identified nuisance parameters. One of them is D. Andrews and Cheng (2011). The authors impose some additional restrictions by assuming that they know the structure of weak identification, namely, they assume that it is known which parameters are potentially weakly identified

and that there is a parameter which defines the strength of identification. This assumption is not directly applicable to our DSGE and Phillips curve examples. Andrews and Cheng (2011) showed that in their case statistics for testing hypotheses with weakly identified nuisance parameters have non-standard asymptotic distributions which depend on the value of the nuisance parameter. Their procedure is based on simulating the asymptotic distribution of the test statistic for different values of the nuisance parameter and taking the "least favorable" among those distributions over a set of relevant nuisance parameter values.

Another example of inference with weakly identified nuisance parameters is given in I. Andrews and Mikusheva (2011). That paper considers a case when concentrating out a weakly identified parameter leads to asymptotically correct inferences, but this result holds only for weakly identified parameters which enter the log-likelihood function linearly.

5 Working with subset of parameters

Suppose we have a single observation $\widehat{\theta}$ from a population $\widehat{\theta} \sim N(\theta_0, \Sigma)$ with an unknown mean θ_0. We wish to test a hypothesis of the form $H_0 : \theta_0 = \theta(\beta)$ for some value of the p-dimensional structural parameter $\beta \in U \subset \mathbb{R}^p$. Our testing procedure suggested in Lemma 1 treats all components of the multi-dimensional vector β in such a way that only the direction of highest curvature affects the value of $\widehat{C}$ and thus influences the critical values. Imagine instead that β can be divided into two sub-sets of parameters $\beta = (\beta_1', \beta_2')'$ in such a way that the curvature corresponding to directions β_1 is high, but the null hypothesis seems to be close to flat in the parameter β_2. Let p_1 be the dimension of β_1, and p_2 the dimension of β_2: $p = p_1 + p_2$. In this section we propose modifications to our testing procedure (suggested in Lemma 1) and pre-test described in Section 4.1 that treat β_1 and β_2 differently. In particular we reduce the critical value of the test due to the low curvature with respect to β_2 while projecting over β_1. The modified procedure may be more efficient if the difference in curvature with respect to parameters β_1 and β_2 is large.

We start with a modification of the pre-test described in Section 4.1. This modification may be thought of as a pre-test for the possibility of concentrating out β_2 in the AR statistic defined in (14). The two competing options are 1) use our robust critical values

based on curvature for the full parameter vector or project over the full parameter vector; or 2) project over β_1, concentrate out β_2 and use $\chi^2_{k-p_2}$ critical values.

The suggested procedure is the following. For any value β_1 consider a p_2-dimensional manifold $S(\beta_1) = \{\Sigma^{-1/2} m(\beta_1, \beta_2), \beta_2 \in \mathbb{R}^{p_2}\}$. For any point $q = \Sigma^{-1/2} m(\beta_1, \beta_2) \in S(\beta_1)$ find the curvature $\kappa_q(S(\beta_1))$. Let

$$\widehat{C} = \left(\min_{\beta_1} \min_{q \in S(\beta_1) \bigcap \widetilde{B}} 1/\kappa_q(S(\beta_1)) \right) \wedge R, \tag{15}$$

where $\widetilde{B}$ is the ball of radius $(1+\sqrt{2})R$ around the point $\widehat{x} = \Sigma^{-1/2}\widehat{\theta}$. One also needs to check that as stated in Lemma 1 the analog of Assumption 1 is satisfied but now we are considering only p_2-dimensional balls in the tangent space to $S(\beta_1)$. If $\widehat{C} > C^*$ it is safe to use the second approach (that is, to concentrate out β_2), while otherwise the first approach should be used. We can likewise adapt the conclusion of Lemma 1 to state that the test which rejects the null if and only if

$$AR = \min_{\beta} (\widehat{\theta} - \theta(\beta))' \Sigma^{-1} (\widehat{\theta} - \theta(\beta)) > F_{1-\alpha}(\widehat{C}, R, k, p_2)$$

controls size.

Curvature of strongly identified parameters. If a set of nuisance parameters is strongly identified in the sense of Stock and Wright (2000), the null hypothesis is asymptotically linear in these parameters as we argued in Section 2.1. Here we show that the curvature corresponding to these parameters is asymptotically of order $O(1/\sqrt{n})$, where n is the sample size.

Consider a sample of size n from some model parameterized by structural parameter β that belongs to some bounded set $U \subseteq \mathbb{R}^p$ and assume that one can estimate the reduced-form parameters θ in a consistent and asymptotically normal way:

$$\sqrt{n}(\hat{\theta} - \theta) \Rightarrow N(0, \Sigma_0).$$

Assume that the relation between structural and reduced-form parameters $\theta = \theta(\beta)$ is fixed (not changing with n), twice continuously differentiable, and that the matrix $\frac{\partial}{\partial \beta} m(\beta)$ has full rank in a neighborhood of β_0, which is the only point in the closure of

U that solves the equation $\theta_0 = \theta(\beta)$. The null hypothesis manifold for sample size n is $S_n = \{\sqrt{n}\Sigma_0^{-1/2} m(\beta), \beta \in U\} \subset \mathbb{R}^k$. The maximal curvature over all points of the manifold S_n is equal to $1/\sqrt{n}$ multiplied by the maximal curvature of the manifold S_1 obtained for sample size 1. This can easily be seen by examining the role of the scale of m in formula (10).

6 Example: A Small-scale DSGE Model

To illustrate the utility of our theoretical results and suggest directions for future research, we apply our approach to a small-scale DSGE model based on Clarida, Gali and Gertler (1999). The (log-linearized) equilibrium conditions for the model are

$$\begin{cases} bE_t\pi_{t+1} + \kappa x_t - \pi_t + \varepsilon_t = 0, \\ -[r_t - E_t\pi_{t+1} - rr_t^*] + E_t x_{t+1} - x_t = 0, \\ \lambda r_{t-1} + (1-\lambda)\phi_\pi \pi_t + (1-\lambda)\phi_x x_t + u_t = r_t, \\ rr_t^* = \rho \Delta a_t, \end{cases} \tag{16}$$

where the exogenous variables (Δa_t and u_t) evolve according to

$$\Delta a_t = \rho \Delta a_{t-1} + \varepsilon_{a,t}; \quad u_t = \delta u_{t-1} + \varepsilon_{u,t};$$
$$(\varepsilon_t, \varepsilon_{a,t}, \varepsilon_{u,t})' \sim iid\, N(0, \Sigma); \Sigma = diag(\sigma^2, \sigma_a^2, \sigma_u^2).$$

Here we assume that a researcher observes data on inflation π_t, the interest rate r_t and some measure of real activity x_t. This model has ten parameters: the discount rate b, the structural parameters κ, ϕ_x, ϕ_π, and λ, and the parameters describing the evolution of the exogenous variables. We calibrate the structural parameters at generally accepted values similar to those used by Mueller (2010): $b = .99$, $\kappa = \frac{(1-\theta)(1+\phi)(1-b\theta)}{\theta} \approx .1717$, $\phi_x = 0.25$, $\phi_\pi = 1.5$ and $\lambda = 0.5$. For the parameters describing the exogenous variables, we choose $\rho = .2$ and $\delta = .2$ to introduce a degree of persistence while maintaining stationarity, and set $\sigma_a = .38$, $\sigma_u = .31$, and $\sigma = 1$. We generate samples of size 300 from this model and then discard the first 100 observations, using only the last 200 observations for the remainder of the analysis. Given well-documented problems with estimating b in many models, from this point forward we calibrate this parameter at its true value and treat the remaining 9 parameters as unknown.

In Andrews and Mikusheva (2011) we documented the poor performance of the classical Maximum Likelihood Estimator for this model and showed that classical confidence

sets are unreliable. In particular, we argued that the model displays behavior consistent with weak identification. Canova and Sala (2009) argued that many DSGE models suffer from weak identification. Simulation evidence in Ruge-Murcia (2010) also shows substantial overrejection for simulated method of moments-based Wald tests in some models. Unfortunately, however, the source and extent of weak identification in DSGE models is not well-understood. Most DSGE models are highly nonlinear in parameters, and are difficult if not impossible to solve analytically.

Several procedures have recently been proposed for testing simple hypotheses, that is hypotheses on the full parameter vector, in DSGE models which may be weakly identified. These tests include the score tests of I. Andrews and Mikusheva (2011) and Qu (2011), the LR statistic proposed by Guerron-Quintana, Inoue and Kilian (2009), and the test of Dufour, Khalaf, and Kichian (2009). Most of these papers suggest the projection method for testing hypotheses on subsets of parameters and confidence set construction, the exception being I. Andrews and Mikusheva (2011) who suggest a procedure for concentrating out strongly identified nuisance parameters. The main problem with the projection method is that it tends to be very conservative, since DSGE models typically have a large number of parameters. On the other hand, concentrating out nuisance parameters is also problematic as it is generally not clear which parameters are weakly identified, and we are unaware of any currently available procedure which would allow us to make this determination.

As outlined in section 2.2.1, to test hypotheses on the structural parameters in DSGE models we can test the implied restriction on the model auto-covariances $\theta(\beta)$. In particular, we let θ consist of the covariance matrix of the observables (x_t, π_t, r_t) and their first auto-covariance, giving us 15 reduced-form parameters. To focus on the problem of weak identification and abstract from the problems which may arise from HAC covariance matrix estimation, we treat the true covariance matrix Σ of our reduced-form parameter estimates as known, and consider AR statistics of the form $AR(\beta) = (\widehat{\theta} - \theta(\beta))' \Sigma^{-1} (\widehat{\theta} - \theta(\beta))$.

To illustrate the application of our approach, we consider the problem of separately testing that each of the structural parameters is equal to its true value (as one one needs to do to construct confidence sets for each parameter individually). For example, to test $H_0 : \kappa = \kappa_0$, we let $\widetilde{\beta}$ contain all the parameters other than κ and consider the AR statistic $AR(\kappa_0) = \min_{\widetilde{\beta}} AR(\kappa_0, \widetilde{\beta})$. As before, the key issue is what critical values to use. The projection method uses the 95th percentile of a χ^2_{15}, which is equal to 25. If

we assume that a q-dimensional sub-vector of $\widetilde{\beta}$ is strongly identified, we can use χ^2_{15-q} critical values instead, which are equal to 14.07 if we take all of $\widetilde{\beta}$ to be strongly identified ($q = 8$). Applying our robust critical values, in contrast, requires no assumption on the strength of identification. As we might expect in a poorly identified model, tests which concentrate out the nuisance parameters do not control size. We simulated tests for each parameter separately, and almost all of them over-reject, though the degree of over-rejection is limited. For example, nominal 5% tests for ρ and σ_a which concentrate out the nuisance parameters have size 9.2% and 9.6% respectively.

For each of the nine parameters, to compute our robust critical values we calculate the curvature of the submanifold of $\{\Sigma^{-\frac{1}{2}}\theta(\beta)\}$ obtained by holding that parameter equal to its null value, intersected with the ball $B_{\sqrt{2}R}(\theta(\beta_0))$ of radius $\sqrt{2}R$ around θ_0, for R the .99 quantile of a χ^2_{15} distribution. We find quite substantial curvature: the manifold implied by $\kappa = \kappa_0$, for example, has a maximal curvature of 1.72 which gives a robust critical value of 23.8. While smaller than the projection-method critical value, this is still quite large.

Not all parameters play an equal role in generating this curvature, however. As noted in I. Andrews and Mikusheva (2011), some parameters seem to be strongly identified, while other seem quite weak. To relate this to curvature, we consider projecting over different subsets of parameters as described in section 5. We find that by projecting over the Taylor-rule parameters ϕ_x and ϕ_π we can reduce the curvature dramatically, suggesting that the group of parameters other than ϕ_x and ϕ_π may be significantly better identified. Hence, to obtain smaller critical values, when testing hypotheses on the structural parameters we project over ϕ_x and ϕ_π.

For each structural parameter, Table 2 reports the robust critical value obtained from this exercise (column 2), together with the simulated size (based on 500 simulations) of nominal 5% tests based on our robust critical values (column 3), and projection-method tests (column 4). As we can see, projection-method based tests (using critical values of 25) are extremely conservative, with simulated size less than or equal to 0.2%. Our robust critical values range from 18.65 to 20.3 for different parameters, and the size of tests using these critical values (ranging from .4% to 1.6%) never exceeds the nominal size. At the same time, these tests are substantially less conservative than the projection method.

Parameter tested	Robust Critical Value	Robust Test Size	Projection Size
ϕ_x	18.72	1.20%	0.00%
ϕ_π	18.65	0.80%	0.00%
λ	19.86	1.60%	0.00%
ρ	19.68	1.40%	0.00%
δ	20.30	1.40%	0.20%
κ	19.51	1.40%	0.20%
σ_a	19.84	1.20%	0.00%
σ_u	19.27	0.40%	0.00%
σ	19.60	1.40%	0.00%

Table 2: Nominal 5% Tests of one-dimensional hypotheses on structural parameters. The first column lists the tested parameter for each row, while the other parameters are treated as nuisance parameters. The statistic is AR minimized over nuisance parameters. Projection method critical values are 25. Robust critical values are based on projecting over ϕ_x and ϕ_π.

7 Appendix with proofs

7.1 Proof of Theorem 1

The proof is based on the following lemma:

Lemma 2 *Assume the curve $\alpha(s) : [0, b] \to D_C \subset \mathbb{R}^k$ is parameterized by arc length and that its curvature $\kappa(s) = \|\ddot{\alpha}(s)\| < \frac{1}{C}$ for all points s. Assume that $\alpha(0) = 0$ and $\dot{\alpha}(0) = v \in$ span$\{e_1, ..., e_p\}$, where $e_1, ..., e_p$ are first p basis vectors. Then the curve $\alpha(s)$ is contained in the set $M_v \cap D_C$, where*

$$M_v = \{x : \langle x, v\rangle^2 + (C - \|x - \langle x, v\rangle v\|)^2 \geq C^2\}. \tag{17}$$

Proof of Lemma 2.

Consider the curve defined by $\beta(s) = \dot{\alpha}(s)$, the first derivative of α. Since the curve α is parameterized by arc length $\|\beta(s)\| = \|\dot{\alpha}(s)\| = 1$ and the new curve β lies on the unit sphere $Sph = \{x \in \mathbb{R}^k : \|x\| = 1\}$, with $\beta(0) = v$. Let $t \leq \frac{\pi}{2}C$ and $t \leq b$. Consider the arc length of the restriction of the curve β to the interval $[0, t]$:

$$length(t) = \int_0^t \|\dot{\beta}(s)\|ds = \int_0^t \|\ddot{\alpha}(s)\|ds = \int_0^t \kappa(s)ds \leq \frac{t}{C}.$$

This implies that the geodesic (a curve of a shortest length) on the sphere Sph connecting $\beta(0)$ and $\beta(t)$ has length less than or equal to $\frac{t}{C}$ or, equivalently, that the angle between

vectors $\beta(0) = v$ and $\beta(t)$ is less than or equal to $\frac{t}{C}$. Hence

$$\langle v, \beta(t)\rangle = \langle v, \dot{\alpha}(t)\rangle \geq \cos(\frac{t}{C}). \tag{18}$$

Since $\alpha(s)$ is parameterized by arc length, from inequality (18) we have:

$$\|\dot{\alpha}(t) - \langle v, \dot{\alpha}(t)\rangle v\| \leq |\sin(\frac{t}{C})|. \tag{19}$$

This, in turn, implies that

$$\|\alpha(t) - \langle v, \alpha(t)\rangle v\| = \|\int_0^t (\dot{\alpha}(s) - \langle v, \dot{\alpha}(s)\rangle v)ds\| \leq$$
$$\leq \int_0^t \|\dot{\alpha}(s) - \langle v, \dot{\alpha}(s)\rangle v\|ds \leq \int_0^t \sin(\frac{s}{C})ds = C - C\cos(\frac{t}{C})$$

Inequality (18) also implies that

$$\langle v, \alpha(t)\rangle \geq \int_0^t \cos(\frac{s}{C})ds = C\sin(\frac{t}{C}). \tag{20}$$

Combing these results yields

$$\langle v, \alpha(t)\rangle^2 + (C - \|\alpha(t) - \langle v, \alpha(t)\rangle v\|)^2 \geq C^2$$

for all $t \leq \frac{\pi}{2}C$. Notice that (20) implies that for $\tau = \frac{\pi}{2}C$ we have $\langle v, \alpha(\tau)\rangle \geq C$ and thus for the first p coordinates of $\alpha(\tau)$, which we denote $\alpha^{(1)}(\tau)$, we have $\|\alpha^{(1)}(\tau)\| \geq C$ so the curve is leaving or has already left the cylinder D_C and thus $b \leq \frac{\pi}{2}C$. This concludes the proof of the lemma. □

Proof of statement (a) of Theorem 1. First, let us show that

$$\bigcup_{\substack{v\in T_0(S)\\ \|v\|=1}} M_v = \{\|x^{(1)}\|^2 + (C - \|x^{(2)}\|)^2 \geq C^2\} = \mathcal{M}, \tag{21}$$

where M_v is defined in (17), $\mathcal{M}$ is defined in (11) and $T_0(S)$ is the tangent space to S at zero and is spanned by first p basis vectors. Indeed, the set on the left hand side consists

of points x for which there exists a vector $v \in span\{e_1, ..., e_p\}, \|v\| = 1$, such that

$$\langle x, v\rangle^2 + (C - \|x - \langle x, v\rangle v\|)^2 \geq C^2. \tag{22}$$

For each x let us find the maximum of the expression on left-hand side of inequality (22) over $v \in T_0(S), \|v\| = 1$:

$$\begin{aligned} &\langle x, v\rangle^2 + (C - \|x - \langle x, v\rangle v\|)^2 = \\ &= \langle x, v\rangle^2 + C^2 + \|x\|^2 - \langle x, v\rangle^2 - 2C\|x - \langle x, v\rangle v\| = \\ &= C^2 + \|x\|^2 - 2C\|x - \langle x, v\rangle v\| \end{aligned}$$

where we used that $\|x - \langle x, v\rangle v\|^2 = \|x\|^2 - \langle x, v\rangle^2$. We see that maximizing the left-hand side of (22) over $v \in span\{e_1, ..., e_p\}, \|v\| = 1$ is equivalent to minimizing $\|x - \langle x, v\rangle v\|$. The minimum is achieved at $v = \frac{1}{\|x^{(1)}\|}(x^{(1)}, 0, ..., 0)$, where $x^{(1)} \in \mathbb{R}^p$ consists of the first p components of x. As a result, the maximum of the left-hand side of (22) equals

$$C^2 + \|x\|^2 - 2C\|x^{(2)}\| = \|x^{(1)}\|^2 + (C - \|x^{(2)}\|)^2.$$

This proves statement (21).

Now assume that the statement (a) of Theorem 1 is incorrect and there exists a point $q \in S_C$ that $q \notin \mathcal{M}$. Take a geodesic line (a curve of the shortest distance lying in S_C) $\alpha(s)$ connecting q and 0 lying in S_C, where such curve exists since S_C is a connected manifold. Parameterize this curve by the arc length. The curve $\alpha(s)$ is geodesic in S if and only if at any point $q = \alpha(t)$ the second derivative $\ddot{\alpha}(t)$ is perpendicular to $T_q(S)$ (see Spivak (1999) for discussion of geodesics, v. 3, p.3). As a result, the curvature of the geodesic α at each point $q = \alpha(t)$ is equal to $\kappa_q(X, S)$ (where $X = \dot{\alpha}(t)$), and thus it is less than $\frac{1}{C}$. Denote the tangent to this curve at 0 by $v \in T_0(S)$. Applying Lemma 2 we obtain that the curve belongs to $M_v \cap D_C$ and thus belongs to $\mathcal{M} \bigcap D_C$. We have arrived at a contradiction. □

Proof of statement (c) of Theorem 1. Let $f(u) = \rho(\xi, N_u)$. We need to find the maximizer of $f(u)$ subject to the constraint $\|u\| = 1$. To differentiate $f(u)$ we use the "envelope theorem" that allows one to differentiate a function which is the optimum of a constrained optimization problem and yields $\frac{df(u)}{du} = \xi^{(2)} - zu$. Hence, the first-order

condition for finding $\widetilde{u}$ implies that u is proportional to $\xi^{(2)}$. The sign is a reflection of the fact that we search for a max rather than a min. $\square$

Proof of statement (b) of Theorem 1. For a given point $\xi \in \mathbb{R}^k$ find the sphere $N_{\widetilde{u}}$ furthest from ξ, $\widetilde{u}$ is described in Theorem 1 (c), and the point $\tau \in N_{\widetilde{u}}$ such that $\rho(\xi, N_{\widetilde{u}}) = \rho(\xi, \tau)$. Consider the $k-p$ dimensional linear space $R_\tau = \{x \in \mathbb{R}^k : x^{(1)} = \tau^{(1)}\}$ that restricts the first p components of x to coincide with the first p components of τ. We will put forward two statements: first, that all points in the intersection of $R_\tau \bigcap \mathcal{M} \bigcap D_C$ are not further from ξ than τ; and second, that this intersection $R_\tau \bigcap \mathcal{M} \bigcap D_C$ contains at least one point from S. Together, these two statements imply that $\rho(\xi, S) \leq \rho(\xi, \tau)$.

The intersection of the three sets $R_\tau \bigcap \mathcal{M} \bigcap D_C$ can be written as follows:

$$R_\tau \bigcap \mathcal{M} \bigcap D_C = \{x = (\tau^{(1)}, x^{(2)}) \in D_C : \|\tau^{(1)}\|^2 + (C - \|x^{(2)}\|)^2 \geq C^2\} = \\ = \left\{x = (\tau^{(1)}, x^{(2)}) : \|x^{(2)}\| \leq C - \sqrt{C^2 - \|\tau^{(1)}\|^2}\right\}.$$

Now let us show that for each $x \in R_\tau \bigcap \mathcal{M} \bigcap D_C$ we have $\rho(\xi, x) \leq \rho(\xi, \tau)$. Indeed, one can solve the constrained maximization problem

$$\rho(\xi, x)^2 = \|\xi^{(1)} - \tau^{(1)}\|^2 + \|\xi^{(2)} - x^{(2)}\|^2 \to \max \text{ s.t. } x \in R_\tau \bigcap \mathcal{M} \bigcap D_C.$$

From the first-order condition for this problem one can see that the maximum is achieved at $x^{(2)}$ proportional to $\xi^{(2)}$, and further inspection reveals that it is achieved at $x = \tau$. Hence, all points lying in the intersection $R_\tau \bigcap \mathcal{M} \bigcap D_C$ have distance to ξ less or equal than $\rho(\xi, N_{\widetilde{u}})$.

To complete the proof we need only show that $R_\tau \bigcap \mathcal{M} \bigcap D_C$ contains at least one point from the manifold S. Recall that from the definition of $\tau \in N_{\widetilde{u}}$ it follows that $\|\tau^{(1)}\| \leq C$. Then Assumption 1 guarantees that the intersection of S_C with R_τ is non-empty, while statement (a) of Theorem 1 implies that $S_C \subseteq \mathcal{M} \bigcap D_C$. $\square$

Proof of statement (d) of Theorem 1. Note that both ξ and $N_{\widetilde{u}}$ belong to the same $p+1$- dimensional linear sub-space $L_{\widetilde{u}} = \{x : x = (x^{(1)}, -z\widetilde{u}), x^{(1)} \in \mathbb{R}^p, z \in \mathbb{R}\}$. Let us restrict our attention to this subspace only. Let $(x^{(1)}, z)$ be the coordinate system in this sub-space, so ξ corresponds to $\widetilde{\xi} = (\xi^{(1)}, \|\xi^{(2)}\|)$, and $N_{\widetilde{u}}$ corresponds to the sphere $N^C = \{x = (x^{(1)}, z) \in \mathbb{R}^{p+1} : \|x^{(1)}\|^2 + (C + z)^2 = C^2\}$. The distance implied by the

distance in $\mathbb{R}^k$ is the usual Euclidean metric, which we denote by $\widetilde{\rho}$. So far, we proved that $\rho(\xi, N_{\widetilde{u}}) = \widetilde{\rho}(\widetilde{\xi}, N^C)$. By invariance of the distance to orthonormal transformations of first p components we have $\widetilde{\rho}(\widetilde{\xi}, N^C) = \widetilde{\rho}(\xi^*, N^C)$, where $\xi^* = (\|\xi^{(1)}\|, 0, ..., 0, \|\xi^{(2)}\|) \in \mathbb{R}^{p+1}$. From this it is easy to see that

$$\rho(\xi, N_{\widetilde{u}}) = \rho_2(\eta, N_2^C),$$

where $\eta = (\|\xi^{(1)}\|, \|\xi^{(2)}\|) \in \mathbb{R}^2$, $N_2^C = \{(z_1, z_2) \in \mathbb{R}^2 : z_1^2 + (C + z_2)^2 = C^2\}$, and ρ_2 is Euclidian distance in $\mathbb{R}^2$. It then follows that if $\xi \sim N(0, I_k)$ then components of η have independent $\sqrt{\chi_p^2}$ and $\sqrt{\chi_{k-p}^2}$ distributions, respectively. $\square$

7.2 Proof of Lemma 1

Proof of Lemma 1. Let $\xi = \Sigma^{-1/2}(\widehat{\theta} - \theta_0) \sim N(0, I_k)$ and $S = \{\Sigma^{-1/2}(\theta - \theta_0), \theta \in H_0\} \subset \mathbb{R}^k$. Let $\psi_C(\xi, R)$ be defined as

$$\psi_C(\xi, R) = \begin{cases} \rho^2(\xi, N_{\widetilde{u}}), & \text{if } \|\xi\| \leq R; \\ \|\xi\|^2, & \text{if } \|\xi\| > R, \end{cases}$$

where $N_{\widetilde{u}} = \{x \in \mathbb{R}^k : x = (x^{(1)}, z\widetilde{u}), x^{(1)} \in \mathbb{R}^p, z \in \mathbb{R}_+, \|x^{(1)}\|^2 + (C - z)^2 = C^2\}, \widetilde{u} = -\frac{1}{\|\xi^{(2)}\|}\xi^{(2)}$. Consider the infeasible test φ which rejects ($\varphi = 1$) if and only if $\psi_C(\xi, R) \geq F_{1-\alpha}(C, R, k, p)$. The size $E\varphi(\xi) = \alpha$, so since $P\{\chi_k^2 \geq R^2\} < \alpha$ we know that φ rejects for all realizations of ξ where $\|\xi\| > R$. This test is infeasible, however, since we do not know the true value of θ_0 and hence cannot calculate ξ. The (feasible) test described in Lemma 1 is

$$\widetilde{\varphi} = \begin{cases} 1, & \text{if } AR \geq F_{1-\alpha}(\widehat{C}, R, k, p); \\ 0, & \text{otherwise.} \end{cases} \tag{23}$$

We claim that $\widetilde{\varphi} \leq \varphi$ almost surely (realization-by-realization). To show that this is the case, assume that $\widetilde{\varphi} = 1$. If at the same time $\|\xi\| > R$ then $\varphi = 1$, so the claim holds. If, on the other hand, $\|\xi\| \leq R$, then the cylinder $\widetilde{D}_R(x_0)$ around $x_0 = \Sigma^{-1/2}\theta_0$ lies inside of ball $\widetilde{B}$, and thus

$$\widehat{C} = \left(\min_{q \in \widetilde{S} \bigcap \widetilde{B}} 1/\kappa_q(\widetilde{S}) \right) \wedge R \leq \left(\min_{q \in \widetilde{S} \bigcap \widetilde{D}_R(x_0)} 1/\kappa_q(\widetilde{S}) \right) \wedge R \leq C.$$

Indeed, to justify the last inequality, consider two cases $R \leq C$ and $R > C$. In the first case $\widehat{C} \leq R \leq C$, in the second case $\widetilde{D}_C \subset \widetilde{D}_R$ and thus $\min_{q \in \widetilde{S} \bigcap \widetilde{D}_R(x_0)} 1/\kappa_q(\widetilde{S}) \leq \min_{q \in \widetilde{S} \bigcap \widetilde{D}_C(x_0)} 1/\kappa_q(\widetilde{S}) \leq C$.

Note that the function $F_{1-\alpha}(c, R, k, p)$ is decreasing in c, and hence $F_{1-\alpha}(C, R, k, p) \leq F_{1-\alpha}(\widehat{C}, R, k, p)$. Further, all the assumptions of Theorem 1 are satisfied so $AR = \rho^2(\xi, S) \leq \rho^2(\xi, N_{\widetilde{u}}) \leq \psi_C(\xi, R)$. Combining these results we obtain that

$$F_{1-\alpha}(C, R, k, p) \leq F_{1-\alpha}(\widehat{C}, R, k, p) \leq AR = \rho^2(\xi, S) \leq \psi_C(\xi, R),$$

and thus $\varphi = 1$. Hence whenever $\widetilde{\varphi} = 1$, we get that $\varphi = 1$ as well, so $\widetilde{\varphi} \leq \varphi$ as we wanted to show, and the size of the feasible test $\widetilde{\varphi}$ is bounded above by α, completing the proof. □

8 References

Anderson, G., and G. Moore (1985): "A Linear Algebraic Procedure for Solving Linear Perfect Foresight Models," *Economics Letters*, 17(3),247-252.

Andrews, D.W.K., and X. Cheng (2011): "Estimation and Inference with Weak, Semi-strong and Strong Identification," *Econometrica, forthcoming.*

Andrews, I. and A. Mikusheva (2011): "Maximum Likelihood Inference in Weakly Identified Models," *unpublished manuscript.*

Blanchard, O.J., and C.M. Kahn (1980): "The Solution of Linear Difference Models under Rational Expectations," *Econometrica*, 48(5),1305-1311.

Canova, F., and L. Sala (2009): "Back to Square One: Identification Issues in DSGE Models," *Journal of Monetary Economics*, 56, 431-449.

Christiano L.J. and M. Eichenbaum (1992): "Current Real-Business-Cycle Theories and Aggregate Labor-Market Fluctuations," *American Economic Review*, 82, 430-450.

Clarida, R., J. Gali, and M. Gertler (1999): "The Science of Monetary Policy: A New Keynesian Perspective," *Journal of Economic Literature*, 37, 1661-1707.

Dufour, J.M., (2003): "Identification, Weak Instruments, and Statistical Inference in Econometrics," *The Canadian Journal of Economics*, 36, 767-808.

Dufour, J.-M., Jasiak, J., (2001): "Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors," *International Economic*

Review 42, 815-843.

Dufour, J.M., L. Khalaf, and M. Kichian (2006): "Inflation Dynamics and the New Keynesian Phillips Curve: An Identification Robust Econometric Analysis," *Journal of Economic Dynamic and Control*, 30, 1707-1727.

Dufour, J.M., L. Khalaf, and M. Kichian (2009): "Structural Multi-Equation Macroeconomic Models: Identification-Robust Estimation and Fit," *Bank of Canada Working Paper*.

Dufour, J.M., and M. Taamouti (2005): "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," *Econometrica*, 73, 1351-1365.

Efron, B. (1975): "Defining the Curvature of a Statistical Problem (with Applications to Second Order Efficiency)," *Annals of Statistics*, 3, 1189-1242.

Gali, J., and M. Gertler (1999): "Inflation Dynamics: A Structural Econometric Analysis," *Journal of Monetary Economics* 44, 195-222.

Guerron-Quintana, P., A. Inoue, and L. Kilian (2009): "Frequentist Inference in Weakly Identified DSGE Models," *unpublished manuscript*.

Guggenberger, P. and R.J. Smith (2005): "Generalized Empirical Likelihood Estimators and Tests under Partial, Weak and Strong Identification," *Econometric Theory*, 21, 667-709.

Guggenberger, P. and R.J. Smith (2008): "Generalized Empirical Likelihood Tests in Time Series Models with Potential Identification Failure," *Journal of Econometrics*, 142(1), 134-161.

Inoue, A. and B. Rossi (2011): "Testing for Weak Identification in Possibly Nonlinear Models," *Journal of Econometrics*, 161(2), 246-261.

Iskrev, N. (2010): "Local Identification in DSGE," *Journal of Monetary Economics*, 57(2), 189-202.

King, R.G. and M.W. Watson (1998): "The Solution of Singular Linear Difference Systems under Rational Expectations," *International Economic Review*, 39(4), 1015-1026.

Kleibergen, F. (2005): "Testing Parameters in GMM Without Assuming that They Are Identified," *Econometrica*, 73, 1103-1123.

Kleibergen, F. (2007): "Generalizing Weak Instrument Robust IV Statistics towards Multiple Parameters, Unrestricted Covariance Matrices and Identification Statistics," *Journal of Econometrics*, 139(1), 181-216.

Kleibergen, F., and S. Mavroeidis (2009a): "Inference on Subsets of Parameters in GMM without Assuming Identification," *unpublished manuscript*.

Kleibergen, F., and S. Mavroeidis (2009b): "Weak Instrument Robust Tests in GMM and

the New Keynesian Phillips Curve," *Journal of Business & Economic Statistics*, 27(3), 293-311.

Kobayashi, S. and K. Nomizu (1969): "*Foundations of Differential Geometry*," *Interscience Publishers.*

Mavroeidis, S. (2005): "Identification Issues in Forward-looking Models Estimated by GMM with an Application to the Phillips Curve," *Journal of Money, Credit and Banking*, 37, 421-449.

Mavroeidis, S. (2010): "Monetary Policy Rules and Macroeconomic Stability: Some New Evidence," *American Economic Review*, 100(1), 491-503.

Mueller, U. (2010): "Measuring Prior Sensitivity and Prior Informativeness in Large Bayesian Models," *unpublished manuscript.*

Nason, J.M. and G.W. Smith (2008): "Identifying the New Keynesian Phillips Curve," *Journal of Applied Econometrics*, 23, 525-551.

Qu, Z. (2011): "Inference and Specification Testing in DSGE Models with Possible Weak Identification," *unpublished manuscript.*

Rotemberg, J., and M. Woodford (1997): "An Optimization Based Econometric Framework for the Evaluation of Monetary Policy," *NBER Macroeconomic Annual*, 12, 297-346.

Ruge-Murcia, F.J. (2007): "Methods to Estimate Dynamic Stochastic General Equilibrium Models," *Journal of Economic Dynamics & Control*, 31, 2599-2636.

Ruge-Murcia, F.J. (2010): "Estimating Nonlinear DSGE Models by the Simulated Method of Moments: With an Application to Business Cycles," *unpublished manuscript.*

Sims, C.A. (2002): "Solving linear rational expectations models," *Computational Economics*, 20, 1-20.

Spivak, M. (1999): "*A Comprehensive Introduction to Differential Geometry*," *Publish or Perish, Texas.*

Stock, J. H., and J.H. Wright (2000): "GMM With Weak Identification," *Econometrica*, 68, 1055-1096.

Stock, J.H., J.H. Wright, and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics*, 20(4), 518-529.

Stock, J.H. and M. Yogo (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg,* D. W. K. Andrews and J. H. Stock, eds., Cambridge: Cambridge University

Press, pp. 80-108.

Wright, J.H. (2003): "Detecting Lack of Identification in GMM," *Econometric Theory*, 19, 322-330.

Yogo, M., (2004): "Estimating the Elasticity of Intertemporal Substitution when Instruments are Weak," *Review of Economics and Statistics* 86, 797-810.