Testing for Weak Identification in Possibly Nonlinear Models

Atsushi Inoue[*] NCSU

Barbara Rossi[†‡] Duke University

December 24, 2010

Abstract

In this paper we propose a chi-square test for identification. Our proposed test statistic is based on the distance between two shrinkage extremum estimators. The two estimators converge in probability to the same limit when identification is strong, and their asymptotic distributions are different when identification is weak. The proposed test is consistent not only for the alternative hypothesis of no identification but also for the alternative of weak identification, which is confirmed by our Monte Carlo results. We apply the proposed technique to test whether the structural parameters of a representative Taylor-rule monetary policy reaction function are identified.

JEL Classification: C12

Keywords: GMM, Shrinkage, Weak Identification

[*]Corresponding Address: Department of Agricultural and Resource Economics, North Carolina State University, Raleigh, NC 27695-8109. Tel: +1 919 515 5969; fax: +1 919 515 1824; Email: atsushi@unity.ncsu.edu

[†]Department of Economics, Duke University, Durham, NC 27708-0097. Email: brossi@econ.duke.edu

[‡]We are grateful to Jonathan Wright for helpful conversations and providing us with his codes used in Wright (2003) and to the editor, associate editor, two anonymous referees, Craig Burnside, Mehmet Caner, Graham Elliott and participants of the first ERID conference for comments. We thank seminar participants at the University of Washington, Seattle, the Federal Reserve Bank of St. Louis, the UNC-NCSU econometrics workshop and the University of Tokyo for helpful comments. This research was supported by National Science Foundation grants SES-1022125 and SES-1022159 and North Carolina Agricultural Research Service Project NC02265.

1 Introduction

The validity of statistical inference in a growing number of macroeconomic models has been questioned in the recent literature. Many of these models are estimated using first order moment conditions and exploiting exogenous instruments, such as in the widely used Generalized Method of Moments (GMM) estimation procedure. As Nelson and Startz (1990a,b) discovered, however, inference is unreliable when the correlation between instruments and endogenous variables is "weak", a situation referred to as the "weak identification" (or "weak instruments") problem. See Canova and Sala (2009), Iskrev (2007) and Ruge-Murcia (2007) for empirical evidence in dynamic stochastic general equilibrium (DSGE) models, Mavroeidis (2010) for the monetary policy rule, Nason and Smith (2005) and Dufour, Khalaf and Kichian (2006) for the new Keynesian Phillips curve, and Yogo (2004) for consumption Euler equations, to name a few. While methods to construct confidence sets that are robust to weak identification have been recently developed, they can be too large to be informative; in addition, applied researchers are often interested in point estimates, in which case their main interest is in whether a model is identified or not.

This paper proposes a new test for identification by testing the null hypothesis of strong identification against the alternative hypothesis of weak (or no) identification. Our proposed test statistic is based on the distance between two bias-corrected shrinkage extremum estimators. Under the null hypothesis of strong identification, the two estimators converge in probability to the same limit and the proposed test statistic has an asymptotic chi-square distribution. Under the alternative hypothesis of weak identification, they converge weakly to different random variables. Our test overcomes two limitations existing in the literature. First, the proposed test is consistent not only for the alternative hypothesis of no identification but also for the alternative of weak identification, whereas existing tests mainly focus on the alternative hypothesis of strict non-identification. Second, our test has the advantage of being applicable to both linear and nonlinear models that may have a large number of parameters, whereas existing tests can only be applied to models with a limited number of parameters and mainly to linear models or non-linear models where the second derivative is independent of the parameter vector.

In the existing literature on identification, identification is often defined in terms of the underlying probability distribution function (see Hsiao, 1983). In many econometric problems, however, true probability measures or likelihood functions are not available to the econometrician, and parameters are estimated by extremum estimators. In this case, we say that parameters are identified if there is a unique minimizer of the estimation objective function. This definition of identification has been extensively used in the econometric literature (see Amemiya, 1985; Gallant and White, 1988; and Newey and McFadden, 1994, for example). We follow this definition of identification in our paper, and refer to this definition of identification as the "identification condition for extremum estimators".[1]

Identification restrictions traditionally take the form of exclusion restrictions (see Hsiao, 1983). In the linear simultaneous equation model, instruments are exogenous if they are excluded from the equation of interest. However, the validity of instruments also requires instruments to be *relevant*. When instruments are only weakly correlated with the endogenous variables, the TSLS estimator is biased towards the probability limit of the OLS estimator and standard inference performs poorly (Bound, Jaeger and Baker, 1995; Nelson and Startz, 1990a,b). To explain the Monte Carlo findings, Staiger and Stock (1997) and Stock and Wright (2000) propose an alternative asymptotic theory in which the correlation is modeled local to zero, and refer to it as "weak identification". Our paper is interested in this concept of identification, and focuses on the relevance condition while maintaining the assumption that the exogeneity conditions hold. In our paper, we test the null hypothesis that this correlation is nonzero and is not local to zero against the alternative that it is local to zero.

A few other papers have considered tests in the presence of weak instruments. In particular, Stock and Yogo (2005) propose to test the null hypothesis that the correlation between endogeneous variables and instruments is local to zero against the alternative that it is not local to zero. Hahn and Hausman (2002) test the null that this correlation is local to zero against the alternative that it is fixed and different from zero, as we do. Our paper is related to these tests, but differs in a crucial way. The advantage of our test relative to that in Stock and Yogo (2005) is that our test does not rely on the Hessian of the objective function whereas the latter test does. Since the Hessian depends on nuisance parameters in nonlinear models, it is unclear how to extend the methods by Stock and Yogo (2005) and Hahn and Hausman (2002) to nonlinear models. Our test can instead be applied to both linear and non-linear models. Wright (2002) proposes a test for the null hypothesis of strong identification by comparing the volume of Wald confidence sets and that of Stock and Wright's (2000) S confidence set. The difference between the two volumes is bounded in probability when the parameters are strongly identified, and diverges to infinity when parameters

[1] A referee suggested to use "Q-identification" to refer to this definition of identification. Although we like the suggestion of the referee, we believe it would be confusing since a large body of the literature uses this definition of identification. This is why we call it instead the "identification condition for extremum estimators".

are weakly identified (because Wald confidence sets are not robust to weak identification whereas the S set is). A potential drawback of this test is that it is not applicable when the number of parameters is more than two. The rank test of Wright (2003) tests the null hypothesis that the relevance condition does not hold against the alternative that it holds. Because his test does not allow for weak identification, the asymptotic null distribution depends on nuisance parameters that cannot be consistently estimated. In fact, our Monte Carlo experiment shows that the rank test of Wright (2003) can suffer from the size distortion when instruments are weak.

There is also a relationship between the tests proposed in this paper and literatures on (i) tests of rank; (ii) reduced rank regression; (iii) tests of overidentification; (iv) tests of no identification; (v) tests of weak identification; and (vi) empirical applications of tests of weak identification. In Section 3.1 we review these literatures in detail and consider a simple linear IV model to illustrate the differences between the existing tests and our test. The advantages of our approach relative to the above mentioned literatures can be summarized as follows. We test the null of strong identification rather than no identification, so that there is no nuisance parameter under the null hypothesis in our setup. Our test allows us to: (i) avoid highly time-consuming searches over the set of all possible parameter configurations that satisfy the null hypothesis of weak identification (as our null hypothesis is strong identification); (ii) have a test with exact size; and (iii) obtain a test that is suitable for highly parameterized nonlinear models, and therefore is especially useful for researchers interested in addressing issues of identification in macroeconomic models.

The idea of shrinkage has been used in the recent literature on many and weak instruments. Carrasco (2008) considers regularization of two-stage least squares estimators in the presence of many instruments. Okui (2007) uses shrinkage in linear simultaneous equations with many instruments and with many weak instruments. While they focus on the estimation problem in linear simultaneous equations, our focus is on testing for identification in possibly nonlinear models.

Monte Carlo simulations confirm that our test has good size and power for reasonable sample sizes. To show the usefulness of the proposed technique, we present an empirical application to the analysis of identification of the parameters of a Taylor rule monetary policy reaction function. We find that the monetary policy parameters were identified in the pre-Volker period, but not in the Volker-Greenspan era.

The rest of the paper is organized as follows: Section 2 presents the assumptions and the theoretical results. Section 3 shows Monte Carlo results using both the Consumption Capital Asset Pricing Models (CCAPM) and the Taylor rule model. Section 4 provides an empirical application addressing the issue of whether the parameters in the U.S. monetary policy reaction function are identified.

Lastly, we mention notational conventions that are used throughout the paper. Let $\nabla_x f(x)$, $\nabla_{xx} f(x)$ and $\nabla_{xxx} f(x)$ denote the gradient vector $(\partial/\partial x)f(x)$, the Hessian matrix $(\partial^2/\partial x\partial x')f(x)$ and the matrix of third derivatives
$(\partial/\partial x')\text{vec}(\nabla_{xx} f(x))$, respectively. When $x = [x_1', x_2']'$, we will sometimes write $f(x)$ as $f(x_1, x_2)$, not $f([x_1', x_2']')$ to simplify the notation. $\|x\|$ is the Euclidean norm of x, $(\sum_{i=1}^n x_i^2)^{1/2}$ when x is an $(n \times 1)$ vector, and $\|A\|$ is the matrix norm, $\max_{\|x\|=1} \|Ax\|$ when A is an $(m \times n)$ matrix. Finally, I_k denotes the $(k \times k)$ identity matrix.

2 Assumptions and Theorems

Consider an extremum estimator $\hat{\theta}_T$ that maximizes some objective function $Q_T(\theta)$,

$$\hat{\theta}_T \;=\; \text{argmax}_{\theta\in\Theta} Q_T(\theta), \tag{1}$$

where $\Theta \subset \Re^k$. (1) includes maximum likelihood, classical minimum distance estimators and generalized method of moments estimators, as discussed in Gallant and White (1988) and Newey and McFadden (1994). A shrinkage estimator coaxes the parameter estimate in some direction by imposing possibly incorrect restrictions,

$$\tilde{\theta}_T \;=\; \text{argmax}_{\theta\in\Theta} \left[Q_T(\theta) - \frac{\lambda_T}{2}\|\theta - \bar{\theta}\|^2 \right], \tag{2}$$

where $\{\lambda_T\}$ is a sequence of positive constants that converges to zero as $T \to \infty$. A well-known shrinkage estimator is a ridge regression estimator with $\bar{\theta} = 0_{k\times 1}$ (Hoerl and Kennard, 1970a,b). We are interested in testing the null hypothesis of strong identification, whose definition is as follows.

Definition (The Null Hypothesis). Under the null hypothesis, the parameters are strongly identified, that is: $\text{plim}_{T\to\infty} Q_T(\theta)$ is uniquely maximized at some $\theta_0 \in \Theta$, where Θ is compact in $\Re^p$.

Suppose $\theta = [\alpha', \beta']'$ where α is possibly weakly identified and β is always strongly identified. Note that it is possible that there are no strongly identified parameters, and our analysis allows for that possibility. Note that empirical researchers do not need to know which parameters are possibly weakly identified and which are strongly identified in order to implement our method in practice. The distinction between α and β is made only for the theoretical derivations. Our objective is to test the null hypothesis that the parameter $\theta_0 = [\alpha_0', \beta_0']'$ is strongly identified against the alternative hypothesis that α_0 is only weakly identified in a sense that we will make precise shortly.

We will impose the following set of assumptions:

Assumptions.

(a) $\Theta = \Theta_A \times \Theta_B$ is non-empty and compact in $\Re^k$ where $\Theta_A \subset \Re^{k_1}$ and $\Theta_B \subset \Re^{k_2}$, $k_1 + k_2 = k$.

(b) $Q_T(\theta)$ is twice continuously differentiable in θ.

(c) Under the null hypothesis H_0, there is a function $Q(\theta)$ such that

  (i) $Q(\theta)$ is twice continuously differentiable, is uniquely maximized at $\theta_0 = [\alpha_0', \beta_0']' \in \text{int}(\Theta)$, and satisfies $\sup_{\theta\in\Theta} |Q_T(\theta) - Q(\theta)| = o_p(1)$;

  (ii) $T^{1/2}[\nabla_\theta Q_T(\cdot) - \nabla_\theta Q(\cdot)] \Rightarrow Z(\cdot)$ holds on Θ, where $\Rightarrow$ denotes weak convergence of random functions on Θ with respect to the sup norm and $Z(\cdot)$ is a zero-mean Gaussian process with covariance kernel $\Sigma(\theta_1, \theta_2) = E(Z(\theta_1)Z(\theta_2)')$ that is positive definite at $\theta_1 = \theta_2 = \theta_0$; and

  (iii) $\nabla_{\theta\theta} Q(\theta_0)$ is non-singular and $\sup_{\theta\in\Theta} \|\nabla_{\theta\theta} Q_T(\theta) - \nabla_{\theta\theta} Q(\theta)\| = O_p(T^{-1/2})$.

(d) Under the alternative hypothesis H_1 :

  (i) There are stochastic processes on Θ, $Q_\alpha(\theta)$, $Q_{\alpha\beta}(\theta)$ and $Q_\beta(\beta)$, such that $\sup_{\theta\in\Theta} \|Q_T(\theta) - T^{-1}Q_\alpha(\theta) - T^{-1/2}Q_{\alpha\beta}(\theta) - Q_\beta(\beta)\| = O_p(T^{-1/2})$, $\sup_{\alpha\in\Theta_A} \|Q_T(\alpha, \beta_0) - T^{-1}Q_\alpha(\alpha, \beta_0) - Q_\beta(\beta_0)\| = o_p(T^{-1})$, and $\sup_{\theta\in\Theta} |Q_\alpha(\theta)|$ is bounded with probability one;

  (ii) There is a stochastic process $G_\alpha(\theta)$ such that $\sup_{\alpha\in\Theta_A} \|T\nabla_\alpha Q_T(\alpha, \beta_0) - G_\alpha(\alpha, \beta_0)\| = o_p(1)$;

  (iii) There are stochastic processes $H_{\alpha\alpha}(\theta)$, $H_{\alpha\beta}(\theta)$ and $H_{\beta\alpha}(\theta)$ such that $\sup_{\theta\in\Theta} \|T\nabla_{\alpha\alpha} Q_T(\theta) - H_{\alpha\alpha}(\theta)\| = o_p(1)$, $\sup_{\theta\in\Theta} \|T^{1/2}\nabla_{\alpha\beta} Q_T(\theta) - H_{\alpha\beta}(\theta)\| = o_p(1)$, and $\sup_{\theta\in\Theta} \|T^{1/2}\nabla_{\beta\alpha} Q_T(\theta) - H_{\alpha\beta}(\theta)\| = o_p(1)$; and

  (iv) $Q_\beta(\beta)$ satisfies Assumption (c) with $Q(\theta)$, $\theta_0 \in int(\Theta)$, $\nabla_\theta Q_T(\theta)$, $\nabla_\theta Q(\theta)$, $Z(\theta)$, $\Sigma(\theta_1, \theta_2)$, $\nabla_{\theta\theta} Q_T(\theta)$ and $\nabla_{\theta\theta} Q(\theta)$ replaced by $Q_\beta(\beta)$, $\beta_0 \in int(\Theta_B)$, $\nabla_\beta Q_T(\theta)$, $\nabla_\beta Q(\beta)$, $Z_\beta(\theta)$, $\Sigma_{\beta\beta}(\beta_1, \beta_2)$, $\nabla_{\beta\beta} Q_T(\theta)$ and $\nabla_{\beta\beta} Q_\beta(\beta)$, respectively, where Z_β is a k_2-dimensional zero-mean Gaussian process with covariance kernel $\Sigma_{\beta\beta}(\beta_1, \beta_2) \equiv E[Z_\beta(\beta_1) Z_\beta(\beta_2)']$.

(e) $\lambda_T = \kappa T^{-1/2}$ for some $\kappa \in (0, \infty)$.

(f) There is a unique $\alpha^* \in \Theta_A$ that maximizes

$$Q_\alpha(\alpha, \beta_0) + Z_\beta(\alpha, \beta_0)' b^*(\alpha) + \frac{1}{2} b^{*\prime}(\alpha) \nabla_{\beta\beta} Q_\beta(\beta_0) b^*(\alpha) \tag{3}$$

where

$$b^*(\alpha) \;=\; -[\nabla_{\beta\beta} Q_\beta(\beta_0)]^{-1} Z_\beta(\alpha, \beta_0). \tag{4}$$

Remarks.

1. Assumptions (b), (c) and (d) are high-level assumptions. Our definition of weak identification in Assumption (d) follows those of Staiger and Stock (1997) and Stock and Wright (2000). α is weakly identified if the part of the objective function that depends on α vanishes (Assumption d.i) and the Hessian of the objective function with respect to α converges to zero at certain rates (Assumption d.iii). Assumption (d) is satisfied in Staiger and Stock's (1997) linear Instrumental Variable (IV) models in which:

$$Q_T(\theta) \;=\; -\frac{1}{T}(y - Y\theta)' X (X'X)^{-1} X'(y - Y\theta), \tag{5}$$

where y and Y are $T \times 1$ and $T \times k$ matrices of endogenous variables and X is a $T \times \ell$ matrix of exogenous variables linked to the regressors via the relationship $Y = X\Pi_0 + V$, with V being a $T \times k$ matrix of error terms.

In their model, our null and alternative hypotheses simplify to

$$H_0 : \text{rank}(\Pi_0) = k \text{ and } H_1 : \Pi_0 = \Pi_T = T^{-1/2} C, \tag{6}$$

where C is an $\ell \times k$ matrix of constants.

2. Assumption (d) is also satisfied in the generalized IV model considered in Stock and Wright (2000) in which

$$Q_T(\theta) \;=\; -\left[\frac{1}{T}\sum_{t=1}^{T} \phi_t(\theta)\right]' \hat{W}_T \left[\frac{1}{T}\sum_{s=1}^{T} \phi_s(\theta)\right], \tag{7}$$

$$Q_\alpha(\theta) \;=\; -m_1(\theta)' W m_1(\theta), \tag{8}$$

$$Q_{\alpha\beta}(\theta) \;=\; -2 m_1(\theta)' W m_2(\beta), \tag{9}$$

$$Q_\beta(\beta) \;=\; -m_2(\beta)' W m_2(\beta), \tag{10}$$

where $\phi_t(\theta)$ is the moment function evaluated at observation t, $E[T^{-1}\sum_{t=1}^{T}\phi_t(\theta)] = m_1(\theta)/\sqrt{T} + m_2(\beta) + o(1)$, $m_1(\theta)$ and $m_2(\beta)$ are some functions, and $\hat{W}_T$ is a weighting matrix that converges to W. See also Guggenberger and Smith (2005) who consider generalized empirical likelihood estimators under assumptions similar to those of Stock and Wright (2000).

3. We can cast our high-level assumptions into the classical minimum distance (CMD) estimation framework. Suppose that

$$\Pi_0 \ = \ g(\theta_0)$$

where Π_0 denotes a vector of reduced-form parameters, θ_0 denotes structural parameters and $g(\cdot)$ maps the structural parameters into the reduced-form parameters. For example, Π_0 is a vector of impulse responses, θ is a vector of structural parameters of a dynamic stochastic general equilibrium (DSGE) model and $g(\cdot)$ is the mapping implied by the DSGE model. The CMD estimator maximizes

$$Q_T(\theta) \ = \ -\left[\hat{\Pi}_T - g(\theta)\right]' \hat{W}_T \left[\hat{\Pi}_T - g(\theta)\right]$$

where $\hat{\Pi}_T$ is a consistent estimator of Π_0 and $\hat{W}_T$ is the weighting matrix. Assumptions (b) and (c) are satisfied under the standard assumptions, such as asymptotic normality of the estimator of the reduced-form parameters and smoothness of the function $g(\cdot)$. Assumption (d) is satisfied if

$$g(\theta) \ = \ g_\beta(\beta) + T^{-1/2} g_\theta(\theta)$$

under the alternative hypothesis.

4. Under the alternative hypothesis, parameters can be all unidentified, i.e., $\alpha = \theta$, $\beta = \emptyset$, $k_1 = k$ and $k_2 = 0$.

5. While our nonlinear framework is general, our assumptions rule out the use of heteroskedasticity autocorrelation consistent (HAC) covariance matrix estimators. Because the HAC covariance matrix estimator is a nonparametric estimator, it converges at rate slower than $T^{1/2}$ and estimators with HAC covariance matrix estimators will violate Assumption (d). Dynamic models based on rational expectations typically imply that Euler residuals and one-period-ahead forecast errors are serially uncorrelated and do not require the use of HAC covariance matrix estimators.

6. The shrinkage parameter, λ_T, determines the harshness of the penalty term. Assumption (e) requires that λ_T converges to zero so that the two objective functions converge in probability to the same limit. As a result, the two estimators converge in probability to the true parameter value under the null hypothesis. Assumption (e) requires that λ_T does not converge to zero too fast, so that the two estimators behave differently under the alternative hypothesis.

7. Existence of a unique maximizer in Assumption (f) only simplifies the asymptotic distribution of the weakly identified parameter, α. The consistency of our proposed test does not necessarily require this assumption, which is made for convenience only. Stock and Wright (2000, p.1062) impose an analogous assumption in their Theorem 1(ii).

In what follows, we will first derive the asymptotic properties of both the extremum estimator and the shrinkage estimator. Under the null hypothesis (strong identification), both estimators are consistent. However, under the alternative hypothesis (weak identification), the extremum estimator does not converge to any constant whereas the shrinkage estimator converges in probability to the value it is shrunk towards. This implies that one cannot construct a consistent test against weak identification using the extremum estimator and is the reason why we focus on shrinkage estimators in this paper.

Theorem 1 (Asymptotic Distributions of Extremum Estimators). Suppose that assumptions (a)–(f) hold.

(a) Under the null hypothesis,

$$T^{\frac{1}{2}}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0_{k\times 1}, [\nabla_{\theta\theta}Q(\theta_0)]^{-1}\Sigma(\theta_0,\theta_0)[\nabla_{\theta\theta}Q(\theta_0)]^{-1}), \tag{11}$$

$$T^{\frac{1}{2}}(\tilde{\theta}_T - \theta_0 - \lambda_T B_T(\theta_0)) \xrightarrow{d} N(0_{k\times 1}, [\nabla_{\theta\theta}Q(\theta_0)]^{-1}\Sigma(\theta_0,\theta_0)[\nabla_{\theta\theta}Q(\theta_0)]^{-1}), \tag{12}$$

where $B_T(\theta_0) = [M_T(\theta_0)]^{-1}(\theta_0 - \bar{\theta})$ and $M_T(\theta) = \nabla_{\theta\theta}Q(\theta) - \lambda_T I_k$.

(b) Under the alternative hypothesis,

$$[\hat{\alpha}_T', T^{1/2}(\hat{\beta}_T - \beta_0)']' \Rightarrow [\alpha^{*\prime}, b^{*\prime}(\alpha^*)]', \tag{13}$$

$$\begin{bmatrix} \lambda_T T(\tilde{\alpha}_T - \bar{\alpha}) \\ T^{\frac{1}{2}}\left(\tilde{\beta}_T - \beta_0 - \lambda_T[\nabla_{\beta\beta}Q_\beta(\beta_0)]^{-1}(\beta_0 - \bar{\beta})\right) \end{bmatrix}$$
$$\Rightarrow \begin{bmatrix} G_\alpha(\bar{\alpha},\beta_0) - H_{\alpha\beta}(\bar{\alpha},\beta_0)(Z_\beta(\bar{\alpha},\beta_0) - \kappa(\beta_0 - \bar{\beta})) \\ -[\nabla_{\beta\beta}Q_\beta(\beta_0)]^{-1}Z_\beta(\bar{\alpha},\beta_0) \end{bmatrix}, \tag{14}$$

where α^* and $b^*(\alpha)$ are defined in (3) and (4), respectively, in Assumption (f), and $\bar{\theta} = [\bar{\alpha}', \bar{\beta}']' \in \Theta$.

Remarks. Equation (11) in part (a) of Theorem 1 is a standard result for extremum estimators and is presented for reference. Equation (12) shows that the shrinkage estimator has a higher-order bias term but has the same asymptotic distribution as the extremum estimator. This is because λ_T converges to zero at rate $T^{-1/2}$. Part (b) shows that the two estimators behave differently in the presence of weakly identified parameters. As Stock and Wright (2000) show for the GMM estimator, the extremum estimator is inconsistent and converges to a random variable. The shrinkage estimator converges in probability to $\bar{\theta}$ because the restriction imposed on the shrinkage estimator constrains the shrinkage estimator in the limit when the parameter is weakly identified.

Consider two extremum estimators,

$$\hat{\theta}_{1T} = \text{argmax}_{\theta\in\Theta} Q_{1T}(\theta), \tag{15}$$

$$\hat{\theta}_{2T} = \text{argmax}_{\theta\in\Theta} Q_{2T}(\theta), \tag{16}$$

and their shrinkage versions,

$$\tilde{\theta}_{1T} = \text{argmax}_{\theta\in\Theta} \left[Q_{1T}(\theta) - \frac{\lambda_T}{2}\|\theta - \bar{\theta}\|^2 \right], \tag{17}$$

$$\tilde{\theta}_{2T} = \text{argmax}_{\theta\in\Theta} \left[Q_{2T}(\theta) - \frac{\lambda_T}{2}\|\theta - \bar{\theta}\|^2 \right]. \tag{18}$$

For example, $\hat{\theta}_{1T}$ and $\hat{\theta}_{2T}$ can be GMM estimators with identity and optimal weighting matrices. Define a test statistic by

$$\hat{R}_T = \hat{d}_T'(\hat{D}_T'\hat{\Sigma}_T(\hat{\theta}_T)\hat{D}_T)^{-1}\hat{d}_T, \tag{19}$$

where

$$\begin{aligned} \hat{d}_T &= T^{\frac{1}{2}}(\tilde{\theta}_{2T} - \tilde{\theta}_{1T} - \lambda_T\hat{B}_{2T} + \lambda_T\hat{B}_{1T}), \\ \hat{D}_T &= \begin{bmatrix} -[\nabla_{\theta\theta}Q_{1T}(\tilde{\theta}_{1T}) - \lambda_T I_k]^{-1} \\ [\nabla_{\theta\theta}Q_{2T}(\tilde{\theta}_{2T}) - \lambda_T I_k]^{-1} \end{bmatrix}, \end{aligned} \tag{20}$$

$$\hat{\Sigma}_T = \begin{bmatrix} \hat{\Sigma}_{11,T} & \hat{\Sigma}_{12,T} \\ \hat{\Sigma}_{21,T} & \hat{\Sigma}_{22,T} \end{bmatrix}, \tag{21}$$

and $\hat{B}_{jT} = [\nabla_{\theta\theta}Q_{jT}(\tilde{\theta}_{j,T}) - \lambda_T I_k]^{-1}(\hat{\theta}_{jT} - \bar{\theta})$ for $j = \{1, 2\}$, $\hat{\Sigma}_T$ is a consistent estimator of the asymptotic covariance matrix of $T^{1/2}[\nabla_\theta Q_{1T}(\theta_0)'\ \nabla_\theta Q_{2T}(\theta_0)']$.

In order to ensure that the test statistic has a well-defined limiting distribution under the null hypothesis and that the test is consistent under the alternative, we make additional assumptions.

Assumptions.

(g) $\alpha_1^* \neq \alpha_2^*$ with probability one where α_1^* and α_2^* are defined in (3) for $\hat{\theta}_{1T}$ and $\hat{\theta}_{2T}$, respectively.

(h) (i) Under the null hypothesis, $\hat{\Sigma}_T$ is a consistent estimator of $\Sigma \equiv AVar\left(T^{\frac{1}{2}}[\nabla_\theta Q_{1T}(\theta_0)' \ \nabla_\theta Q_{2T}(\theta_0)']\right)$, and $D'\Sigma D$ is non-singular, where

$$D = \begin{bmatrix} -[\nabla_{\theta\theta} Q_1(\theta_0)]^{-1} \\ [\nabla_{\theta\theta} Q_2(\theta_0)]^{-1} \end{bmatrix} \tag{22}$$

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \tag{23}$$

(ii) Under the alternative hypothesis, there are random matrices Σ_{11}^*, Σ_{12}^*, Σ_{21}^* and Σ_{22}^* such that

$$\begin{bmatrix} T^{\frac{1}{2}} I_{k_1} & 0_{k_1 \times k_2} \\ 0_{k_2 \times k_1} & I_{k_2} \end{bmatrix} \hat{\Sigma}_{ij,T} \begin{bmatrix} T^{\frac{1}{2}} I_{k_1} & 0_{k_1 \times k_2} \\ 0_{k_2 \times k_1} & I_{k_2} \end{bmatrix} \Rightarrow \Sigma_{ij}^* \tag{24}$$

for $i, j = 1, 2$.

Remarks.

1. Assumption (g) requires that the two extremum estimators converge to different random variables when the parameters are weakly identified. Consider a linear simultaneous equation model with two endogenous variables, for example. Let $N(\mu, \Sigma)$ denote a normally distributed random vector with mean μ and covariance matrix Σ. Then the GMM estimator with the identity weighting matrix converges weakly to the random variable that maximizes a non-central χ^2 random function of $\alpha \in \Theta_A$:

$$\begin{aligned} &N\left(E(z_i z_i')C(\alpha - \alpha_0), E(\varepsilon_i - (\alpha - \alpha_0)\eta_i)^2 E(z_i z_i')\right)' \\ &\times N\left(E(z_i z_i')C(\alpha - \alpha_0), E(\varepsilon_i - (\alpha - \alpha_0)\eta_i)^2 E(z_i z_i')\right), \end{aligned} \tag{25}$$

where z_i is a $l \times 1$ vector of instruments, C is a $l \times 1$ vector of Pitman drift parameters such that $\Pi = T^{-1/2}C$, ε_i is the disturbance term in the structural equation, and η_i is the disturbance term of the reduced form equation for the endogenous variable included on the right hand side of the structural equation. The two-stage least squares estimator converges weakly to the random variable that maximizes another non-central χ^2 random function of $\alpha \in \Theta_A$:

$$N\left(E(z_i z_i')^{-\frac{1}{2}} C(\alpha-\alpha_0), E(\varepsilon_i-(\alpha-\alpha_0)\eta_i)^2 I_l\right)' \times N\left(E(z_i z_i')^{-\frac{1}{2}} C(\alpha-\alpha_0), E(\varepsilon_i-(\alpha-\alpha_0)\eta_i)^2) I_l\right). \tag{26}$$

Unless the instruments are orthonormal, i.e., $E(z_i z_i') = cI_l$ for some $c > 0$, α^* and α^{**} are different in general and Assumption (g) is satisfied when parameters are weakly identified.

Corollary 4 of Stock and Wright (2000, p.1067) also shows that different weighting matrices lead to different limits of GMM estimators.

2. Assumption (h.i) requires that the two extremum estimators $\hat{\theta}_{1T}$ and $\hat{\theta}_{2T}$ have different asymptotic covariance matrices. For just identified linear regression models, OLS and GLS estimators have different asymptotic covariance matrices in general. In general, however, the assumption is not satisfied for just-identified moment restriction models. For over-identified models, this assumption is likely to be satisfied if the two estimators use different weighting matrices. For example:

Weighting matrix 1	Weighting matrix 2	Assumption (h.i) is satisfied if:
Identity matrix	The inverse of cross product of instruments	Instruments are non-orthogonal
The inverse of cross product of instruments	Optimal weighting matrix	Conditional hetero-skedasticity is present

Similar arguments apply to classical minimum distance estimators. When reduced form parameters, such as the parameters of state space models and impulse responses, are functions of structural parameters, the structural parameters can be estimated from reduced form estimates via minimum distance. Two suitable estimators can be obtained by choosing different minimum distance estimators.

3. In IV and GMM estimation, one can achieve Assumptions (g) and (h.i) by adding a relevant instrument. For example, if $\hat{\theta}_{2T}$ is an IV/GMM estimator based on Z_1, then $\hat{\theta}_{1T}$ is an IV/GMM estimator based on Z_1 and Z_2, where Z_2 is a set of relevant instruments.[2] In empirical macroeconomics, we generally have a plenty of candidates for Z_2, such as lagged values of Z_1.

[2]We thank Don Andrews for pointing this out.

4. As an example of $\hat{\Sigma}_T$, consider a linear IV model. Let

$$\begin{aligned} \hat{\theta}_{1,T} &= (Y'X\hat{W}_TX'Y)^{-1}Y'X\hat{W}_TX'y, & (27) \\ \hat{\theta}_{2,T} &= (Y'XX'Y)^{-1}Y'XX'y, & (28) \end{aligned}$$

where $\hat{W}_T = (1/T)\sum_{i=1}^{T}(y_i - \hat{\theta}_{2,T})^2 X_i X_i'$, X_i', y_i and Y_i' are the ith row of X, y and Y, respectively, and the rest of the notation follows the notation in Remark 1 on Assumptions (a)–(f). Then $\hat{\Sigma}_T$ is an estimate of the covariance matrix of

$$\left[\begin{array}{c} Y'X\hat{W}_TX_i(y_i - \hat{\theta}_{1,T}'X_i) \\ Y'XX_i(y_i - \hat{\theta}_{2,T}'X_i) \end{array} \right]. \tag{29}$$

5. Another example of $\hat{\Sigma}_T$ is for the GMM estimator in the second remark on Assumptions (a)–(f). Let $\hat{\theta}_{1,T}$ and $\hat{\theta}_{2,T}$ be the GMM estimators with weighting matrices $[(1/T)\sum_{t=1}^{T}\phi_t(\hat{\theta}_{2,T})\phi_t(\hat{\theta}_{2,T})']^{-1}$ and I_k. Then $\hat{\Sigma}_T$ is an estimate of the covariance matrix of

$$\left[\begin{array}{c} \sum_{s=1}^{T} D_\theta\phi_s(\hat{\theta}_{1,T})\hat{W}_T\phi_t(\hat{\theta}_{1,T}) \\ \sum_{s=1}^{T} D_\theta\phi_s(\hat{\theta}_{1,T})'\hat{W}_T\phi_t(\hat{\theta}_{2,T}) \end{array} \right], \tag{30}$$

where $D_\theta\phi_s(\theta) = [\nabla_\theta\phi_{s,1}(\theta)\ \nabla_\theta\phi_{s,2}(\theta)\ \cdots\ \nabla_\theta\phi_{s,l}(\theta)]'$ is the Jacobian matrix of $\phi_s(\theta)$ and $l = \dim(\phi_s(\theta))$.

Our main result is the asymptotic distribution of $\hat{R}_T$. We state it formally in the following Theorem.

Theorem 2 (Asymptotic Properties of the Proposed Test Statistic). Suppose that Assumptions (a)–(f) hold for $Q_{1T}(\theta)$ and $Q_{2T}(\theta)$ with common θ_0 as well as Assumptions (g) and (h).

(a) If the null hypothesis H_0 is true,

$$\hat{R}_T \xrightarrow{d} \chi_k^2. \tag{31}$$

(b) If the alternative hypothesis H_1 is true and if $M_1'\Sigma_{11}^*M_1 - M_1'\Sigma_{12}^*M_2 - M_2'\Sigma_{21}^*M_1 + M_2'\Sigma_{22}^*M_2$ is non-singular,

$$\frac{1}{T}\hat{R}_T \Rightarrow \kappa^2 \left[\begin{array}{c} \alpha_2^* - \alpha_1^* \\ 0_{k_2\times 1} \end{array} \right]' (M_1'\Sigma_{11}^*M_1 - M_1'\Sigma_{12}^*M_2 - M_2'\Sigma_{21}^*M_1 + M_2'\Sigma_{22}^*M_2)^{-1} \left[\begin{array}{c} \alpha_2^* - \alpha_1^* \\ 0_{k_2\times 1} \end{array} \right]. \tag{32}$$

where

$$\begin{aligned} M_1 &= \left(\begin{array}{cc} -I_{k_1} & 0_{k_1\times k_2} \\ (\nabla_{\beta\beta}Q_{1,\beta}(\beta_0))^{-1}H_{1,\beta\alpha}(\bar{\alpha},\beta_0) & (\nabla_{\beta\beta}Q_{1,\beta}(\beta_0))^{-1} \end{array} \right), \\ M_2 &= \left(\begin{array}{cc} -I_{k_1} & 0_{k_1\times k_2} \\ (\nabla_{\beta\beta}Q_{2,\beta}(\beta_0))^{-1}H_{2,\beta\alpha}(\bar{\alpha},\beta_0) & (\nabla_{\beta\beta}Q_{2,\beta}(\beta_0))^{-1} \end{array} \right). \end{aligned}$$

If $M_1'\Sigma_{11}^*M_1 - M_1'\Sigma_{12}^*M_2 - M_2'\Sigma_{21}^*M_1 + M_2'\Sigma_{22}^*M_2$ is singular,

$$\frac{1}{T}\hat{R}_T \Rightarrow \infty. \tag{33}$$

Remarks.

1. Theorem 2 shows that one can use central χ^2 critical values to test the null hypothesis of strong identification. This is because there are no nuisance parameters under the null hypothesis.

2. Theorem 2(b) shows that if we construct the test statistic using two non-shrinkage extremum estimators with $\kappa = 0$ the test will be inconsistent. Because the standard extremum estimator is inconsistent under weak identification, $\sqrt{T}(\hat{\theta}_{1T} - \hat{\theta}_{2T})$ diverges at rate $T^{1/2}$. Under the alternative hypothesis of weak identification, however, the asymptotic covariance estimator of $\sqrt{T}(\hat{\theta}_1 - \hat{\theta}_2)$ diverges at rate T by Assumption (d). Therefore the test statistic based on the two extremum estimators with $\kappa = 0$ will be bounded in probability under the alternative hypothesis and thus the test will be inconsistent.[3]

3. Theorem 2(b) shows that the test rejects the null hypothesis with probability approaching one whether parameters are not identified at all or only weakly identified.

4. Theorem 2(b) implies that the power is increasing in κ. That is, the test is more powerful the larger λ_T is. There is a size-power trade-off, however. In general, the type I error of the test is bigger for larger values of λ_T, because there is some approximation error of order $O_p(\lambda_T)$.[4] We will discuss the choice of λ_T in the next section.

3 Literature Review and Local Power Analysis

In this section we provide a discussion of how our paper is related to the literatures on tests of weak identification, tests of overidentification, tests of rank, the reduced rank regression, and the local alternative hypotheses of rank condition. The section also provides some intuition regarding our test in a simplified setup and show that the proposed test has nontrivial asymptotic local power in a simple linear IV model.

[3] In linear IV models, Hahn, Ham and Moon (2010) similarly show that the conventional Hausman (1978) test is invalid when instruments are weak.

[4] See equation (65). When multiplied by $T^{1/2}$ there is error of order $O_p(\lambda_T)$.

3.1 Literature Review

Identification is quite commonly defined as follows: The probability measures P_θ and $P_{\theta'}$ are observationally equivalent if $P_\theta = P_{\theta'}$ (see Definition 2.2 of Hsiao, 1983, p.226, for example). When there are no two probability measures that are observationally equivalent, we say that the true probability measure is identifiable and the population likelihood function achieves its maximum at a unique value (see Lemma 5.35 of van der Vaart, 1998, p.62). In many econometric problems, however, the true probability measures or likelihood functions are not available to the econometrician, and parameters are estimated by extremum estimators that maximize or minimize some estimation objective function, e.g. GMM. In this case, we say that parameters are identified if there is a unique minimizer of the objective function (see Amemiya ,1985, p.106, Gallant and White, 1988, p.19, and Newey and McFadden, 1994, p.2121, for example). We focus on this identification condition for extremum estimators.

The following simple example illustrates how various rank conditions for identification are related.

Example

Suppose that data are generated by

$$\begin{bmatrix} 1 & -\beta_0 \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} y_i \\ Y_i \end{bmatrix} = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} z_{1i} \\ z_{2i} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix}, \tag{34}$$

where $[\varepsilon_{1i} \;\; \varepsilon_{2i}]' \overset{iid}{\sim} (0_{2\times 1}, \Sigma)$ is uncorrelated with $Z_i \equiv [z_{1i} \;\; z_{2i}]'$, Σ is a 2×2 positive definite matrix and $E(Z_i Z_i')$ is full rank and diagonal, $i = 1, ..., N$, where N is the total sample size. The econometrician estimates:

$$y_i = \beta_0 Y_i + u_i, \tag{35}$$

where $u_i = \gamma_{11} z_{1i} + \gamma_{12} z_{2i} + \varepsilon_{1i} = Z_i'\Gamma_1 + \varepsilon_{1i}$ and $\Gamma_1 \equiv [\gamma_{11}, \gamma_{12}]'$, by either OLS or 2SLS where instrumental variables z_{1i} and z_{2i} are excluded from the structural equation. Note that (34) implies the reduced form equations:

$$\begin{aligned} \begin{bmatrix} y_i \\ Y_i \end{bmatrix} &= \frac{1}{\Delta} \begin{bmatrix} \gamma_{11} b_{22} + \beta_0 \gamma_{21} & \gamma_{12} b_{22} + \beta_0 \gamma_{22} \\ \gamma_{21} - \gamma_{11} b_{21} & \gamma_{22} - b_{21} \gamma_{12} \end{bmatrix} \begin{bmatrix} z_{1i} \\ z_{2i} \end{bmatrix} + \frac{1}{\Delta} \begin{bmatrix} b_{22}\varepsilon_{1i} + \beta_0 \varepsilon_{2i} \\ \varepsilon_{2i} - b_{21}\varepsilon_{1i} \end{bmatrix} \\ &= A z_i + w_i \end{aligned} \tag{36}$$

provided $\Delta \equiv b_{22} + \beta_0 b_{21} \neq 0$, where $A \equiv \frac{1}{\Delta} \begin{bmatrix} \gamma_{11} b_{22} + \beta_0 \gamma_{21} & \gamma_{12} b_{22} + \beta_0 \gamma_{22} \\ \gamma_{21} - \gamma_{11} b_{21} & \gamma_{22} - b_{21} \gamma_{12} \end{bmatrix}$. Let $w_i \equiv \frac{1}{\Delta} \begin{bmatrix} b_{22}\varepsilon_{1i} + \beta_0 \varepsilon_{2i} \\ \varepsilon_{2i} - b_{21}\varepsilon_{1i} \end{bmatrix} = \begin{bmatrix} w_{1i} \\ w_{2i} \end{bmatrix}$ and $Var(w_i) = \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12} & \Omega_{22} \end{bmatrix}$.

To simplify our discussion, we focus on the case in which the two instruments are excluded from the structural equation. Cases in which one of the exogenous variables is included in the structural equation can be analyzed in an analogous fashion. The 2SLS maximizes the population objective function

$$Q(\beta) \;=\; E[(y_i - \beta' Y_i)Z_i'][E(Z_i Z_i')]^{-1}E[Z_i(y_i - \beta' Y_i)]. \tag{37}$$

The following two conditions ensure that the 2SLS estimator is consistent and asymptotically normal:

(A) the instruments are exogenous, that is, they are uncorrelated to the disturbance term in the structural equation; this requires that $E\left(Z_i u_i\right) = 0$:

$$\begin{aligned} \text{Validity (exogeneity) condition} \quad &: \quad E\left(Z_i u_i\right) = 0 \\ (\text{i.e. } \gamma_{11} \quad &= \quad \gamma_{12} = 0); \text{ and} \end{aligned} \tag{38}$$

(B) the instruments are relevant, $rank[E(Z_i Y_i)] = 1$ (see Assumptions 3.3 and 3.4 of Hayashi, 2000, p.198 and p.200, respectively, for example). Note that $E(Z_i Y_i) = \frac{1}{\Delta}E(Z_i Z_i')\Pi$, where $\Pi \equiv \begin{pmatrix} \gamma_{21} - \gamma_{11} b_{21} \\ \gamma_{22} - b_{21}\gamma_{12} \end{pmatrix}$, thus $rank[E(Z_i Y_i)] = rank\,(\Pi)$, since $E\left(Z_i Z_i'\right)$ is full rank. Clearly, $rank\,(\Pi) = 1$ if either $\gamma_{21} - \gamma_{11} b_{21} \neq 0$ or $\gamma_{22} - b_{21}\gamma_{12} \neq 0$. Thus, the relevance condition is:

$$\begin{aligned} \text{Relevance condition:} \qquad & rank\,(\Pi) = 1 \text{ or } rank[E(Z_i Y_i)] = 1 \\ \text{i.e. } \gamma_{21} - \gamma_{11} b_{21} \quad &\neq \quad 0 \text{ or } \gamma_{22} - b_{21}\gamma_{12} \neq 0. \end{aligned} \tag{39}$$

Note that, since the words identification and rank conditions have been used to mean different conditions in the literature, we refer to condition (i) as the validity condition and to condition (ii) as the relevance condition.

When the validity condition and the relevance condition are both satisfied, then the 2SLS population objective function (37) has a unique minimum at β_0 and the identification condition for extremum estimators is satisfied. If, in addition, the disturbance terms are Gaussian, these two conditions also imply identification of the true probability measure (which is the definition of identification discussed in the chapter by Hsiao, 1983). We have the following cases:

(a) If the validity conditions jointly hold for both instruments and the relevance condition holds then

$$rank(A) \;=\; rank\left(\begin{bmatrix} \beta_0\gamma_{21} & \beta_0\gamma_{22} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}\right) \;=\; 1. \tag{40}$$

(b) If the validity condition fails ($\gamma_{11} \neq 0$ or $\gamma_{12} \neq 0$) but the relevance condition is satisfied ($\gamma_{21} - \gamma_{11} b_{21} \neq 0$ or $\gamma_{22} - b_{21}\gamma_{12} \neq 0$), then

$$rank(A) \; = \; 2, \tag{41}$$

because $|A| = (\gamma_{11}\gamma_{22} - \gamma_{12}\gamma_{21})(b_{22} + \beta_0 b_{21}) = (\gamma_{11}\gamma_{22} - \gamma_{12}\gamma_{21})\Delta \neq 0$, provided that $\left[\begin{array}{cc} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{array}\right]$ is full rank.

(c) If the validity conditions are satisfied ($\gamma_{11} = \gamma_{12} = 0$) but the relevance condition fails for both instruments ($\gamma_{21} - \gamma_{11} b_{21} = \gamma_{22} - \gamma_{12} b_{21} = 0$), then

$$rank(A) = \; 0. \tag{42}$$

Thus the rank of the (2×2) matrix A is 1 if and only if the validity and relevance conditions are both satisfied.

Here below we discuss in detail the relationship between our paper and: (i) tests of rank; (ii) reduced rank regression; (iii) tests of overidentification; (iv) tests of no identification; and (v) tests of weak identification, paying special attention to the alternative hypotheses of the rank condition. (vi) Finally, we discuss why it is important to focus on tests for weak identification as the alternative hypothesis instead of under-identification by reviewing many papers that recently have empirically encountered such problem.

(i) Our paper is related to the literature of tests of rank of a matrix – see the survey by Anderson (1984), and Cragg and Donald (1996, 1997), Robin and Smith (2000), Gill and Lewbel (1992), Kleibergen and Paap (2006) for recent contributions.

(ii) Note that when $rank(A) < 2$, the reduced form of simple example (36) is a reduced rank regression. The technique of reduced rank regression was introduced by Anderson and Rubin (1949) and Anderson (1951). There exist several applications of tests of rank and reduced rank regressions. In a recent paper, Anderson and Kunitomo (2009) develop tests on coefficients in reduced rank regressions. In our example, their test simplifies to testing:

$$A \left[\begin{array}{c} 1 \\ -\beta_0 \end{array}\right] \; = \; \left[\begin{array}{c} 0 \\ 0 \end{array}\right] \;\; \text{for the null parameter value } \beta_0 \tag{43}$$

so that A has rank 1. Anderson and Kunitomo (2009, Section 4.2) show that their test is robust to weak instruments; however, they require $E(Z_i Y_i) = CN^{-\delta}$,, where $0 < \delta < 1/2$, which is a slower rate than that in our paper.

(iii) Anderson and Rubin (1949, 1950) develop tests of overidentifying restrictions, and Anderson and Kunitomo (1992, 1994) propose tests of block identifiability. These papers focus on testing whether the validity conditions (38) hold and the maintained hypothesis for these tests is that the relevance conditions (39) are satisfied; a rejection of tests implies that some of the validity conditions are not satisfied.[5] In our example, they test the null hypothesis that

$$A\begin{bmatrix} 1 \\ -\beta_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \text{ for some } \beta_0 \tag{44}$$

which boils down to the null (40), against the alternative (41). [6] See also Sargan (1958), Durbin (1959), Hausman (1978), Wu (1973), and Hansen (1982) for tests of overidentifying restrictions and the tests of Newey (1985) and Eichenbaum, Hansen and Singleton (1985) for tests of a subset of such validity conditions. The maintained hypothesis for these tests is that the relevance conditions are satisfied. Even if the validity condition fails, if the relevance condition is not satisfied, these tests may not be consistent.

(iv) Koopmans and Hood (1953) and Wright (2003) propose tests for no identification (42) against the alternative (40). These papers focus on testing the null hypothesis that the relevance conditions do not hold against the alternative that they hold and the maintained hypothesis for these tests is that the validity conditions are satisfied. The test of Wright (2003) tests the rank of the lower 1×2 submatrix of the above A matrix, Π, for example. In their survey, Stock et al. (2002) describe methodologies to detect whether instruments are relevant or not. Among such methods, there is the methodology by Cragg and Donald (1993). Cragg and Donald (1993) propose a rank test on Π to test the null that the instruments are not relevant against the alternative that they are relevant; their test, however, is not capable of determining whether the instruments are "sufficiently strong" so that standard inference is reliable. This is the main problem we are interested in.

(v) Our paper focuses on the relevance condition (39) and maintains the assumption that the exogeneity condition (38) holds. In particular, we focus on the empirically relevant problem, corroborated by Monte Carlo evidence, that even if the relevance condition is technically satisfied in population, if the correlation between instruments and endogenous variables are "weak", the standard asymptotic approximation performs poorly (Bound, Jaeger and Baker, 1995; Nelson and

[5] In addition to tests for identifiability, Anderson and Kunitomo (1992, 1994) also discuss tests for predeterminedness, that is testing the null hypothesis that $cov\,(u_t, w_{2i}) = 0$ against the alternative that $cov\,(u_t, w_{2i}) \neq 0$.

[6] Note that the null parameter value is specified in (43) whereas the value of β_0 is unspecified in (44).

Startz, 1990a,b). To explain the Monte Carlo findings, Staiger and Stock (1997) and Stock and Wright (2000) propose an alternative asymptotic theory in which the correlation is modeled local to zero:

$$E(Z_i Y_i) = C N^{-1} \tag{45}$$

for some 2×1 vector C. Anderson and Kunitomo (2009) also develop tests on coefficients of one structural equation in a set of simultaneous equations that are robust to a particular case of weak instruments, where the rate is slower than the one we consider.

In our paper, we test the null hypothesis that $E(Z_i Y_i)$ has rank 1 (so that (40) is satisfied under the maintained validity assumption) and is *not* local to zero against the alternative that it is local to zero, (45). In this example, $\gamma_{21} - \gamma_{11} b_{21} = C_1 N^{-1/2}$ and $\gamma_{22} - b_{21}\gamma_{12} = C_2 N^{-1/2}$ for some C_1 and C_2. Note that our alternative hypothesis includes the case of no identification considered in the tests in (iv) as a special case with $C_1 = C_2 = 0$. A few papers have considered tests for relevance in the presence of weak instruments: Stock and Yogo (2005) propose to test the null hypothesis (45) against the alternative (40). Hahn and Hausman (2002) test the null (41) against the alternative (45), as we do. Our paper is related to these tests, but differs in a fundamental way. The advantage of our test is that it does not rely on the Hessian of the objective function whereas the test proposed by Stock and Yogo (2005) does. Since the Hessian depends on nuisance parameters in nonlinear models, it is unclear how to extend these methods to nonlinear models. Our test can instead be applied to both linear and non-linear models. In addition, the asymptotic null distributions of the existing tests for testing the null (42) against the alternative (40) depend on nuisance parameters since C_1 and C_2 cannot be consistently estimated, and thus are not robust to weak identification. In fact, our Monte Carlo experiment shows that the rank test of Wright (2003) can suffer from the size distortion when instruments are weak.

(vi) The empirical literature (in particular papers estimating first order conditions in macroeconomics and finance) offers several examples of studies concerned about the weak instrument problem. These studies usually test the strength of instruments by using first stage tests. For example, the Stock and Yogo's (2005) first stage test for weak instruments has been used by Consolo and Favero (2009) to estimate monetary policy functions; Yogo (2004) to estimate the elasticity of intertemporal substitution estimated in the context of an Euler equation involving consumption growth and returns on wealth; Krause, Lopez-Salido and Lubik (2008) for inflation dynamics; Shapiro (2008) to estimate the New Keynesian Phillips Curve using a new proxy for the real marginal cost term; Fuhrer and Rudebusch (2004) for the Euler equation for output. The same methodology is also popular as a robustness check in cross-section studies that use IV estimation. For example, Crowe (2010) uses it to estimate the relationship between IT adoption and the quality of private sector forecasts, Faria and Montesinos (2009) to estimate the relation between growth, income level and freedom indexes; and Alcala and Ciccone (2004) for the link between trade and productivity. Other papers use a first-stage test of F-statistic check with a critical value of 10. See, for example, Park and Kang (2008) on the relationship between education and health; Acemoglu and Johnson (2007) to estimate the effect of life expectancy on economic performance; Doyle (2007) to measure the effects of foster care on children outcomes; DeJuan and Seater (2007) to test the Permanent Income Hypothesis; Temple and Wossmann (2006) in cross country growth regressions; Wossmann and West (2006) for the effects of class-size in school systems; Ait Sahalia, Parker and Yogo (2004) for estimates of the equity premium.

3.2 Local Power Analysis

In this section we provide some intuition regarding our test in a simplified setup and show that the proposed test has nontrivial asymptotic local power in a simple linear IV model. First we will define more general null and alternative hypotheses, and then we find the probability limits and asymptotic distributions of the shrinkage estimator under these alternatives and local alternatives. The general null and alternative hypotheses allow us to derive analytic local power results. These results include Theorem 2 as a special case.

Consider the IV model in remark 1 on page 8 with $k = \ell = 1$:

$$\begin{aligned} y &= Y\theta_0 + U, \\ Y &= X_1\Pi_{1,T} + X_2\Pi_{2,T} + V. \end{aligned}$$

Let

$$d \equiv \lim_{T\to\infty} ln(\Pi_{j,T})/ln(T) \text{ for j=1,2.}$$

The null and alternative hypotheses considered in Section 2 can be written as $H_0 : d = 0$ and $H_1 : d = -1/2$, respectively. Now suppose that $\Pi_{j,T} = c_j T^d$ with $c_j \neq 0$ and define a new null hypothesis by

$$\overline{H}_0 : \ d \in \left(-\frac{1}{4}, 0\right]. \tag{46}$$

and an alternative hypothesis by

$$\overline{H}_1 : \ d \in \left[-\frac{1}{2}, -\frac{1}{4}\right). \tag{47}$$

We assume that X_1 and X_2 are independent and satisfy $c_1^2/E(x_{1,i}^2) \neq c_2^2/E(x_{2,i}^2)$ and that $E(u_i|x_{1,i}, x_{2,i}) = \sigma^2$ where $x_{j,i}$ and u_i denote the i-th row of X_j and U, respectively, where $i = 1, 2, ...T$. Let $\hat{\theta}_{j,T}$ and $\tilde{\theta}_{j,T}$ denote the 2SLS and shrinkage estimator based on X_j for $j = 1, 2$. That is, we consider two IV estimators that use different instruments: $\hat{\theta}_{j,T}$ and $\tilde{\theta}_{j,T}$ denote the IV and shrinkage estimators which use instruments X_j and which maximize objective functions $Q_{j,T}(\theta)$ and $Q_{j,T}(\theta) - \lambda_T \theta^2$, respectively.

First we will consider the probability limit of the shrinkage estimator under the null and alternative hypotheses. Let θ_0 denote the true parameter value. Because the objective function for the shrinkage estimator $\tilde{\theta}_{j,T}$ is

$$Q_{j,T}(\theta) - \frac{1}{2}\lambda_T \left(\theta - \bar{\theta}\right)^2 = \begin{cases} -c_j^2 E(x_{j,i}^2)(\theta - \theta_0)^2 T^{2d} + o_p(T^{2d}), \\ \quad \text{if } d \in \left(-\frac{1}{4}, 0\right], \\ -c_j^2 E(x_{j,i}^2)(\theta - \theta_0)^2 T^{-1/2} - \frac{1}{2}\lambda_T(\theta - \bar{\theta})^2 + o_p(T^{-\frac{1}{2}}) \\ \quad \text{if } d = -\frac{1}{4}, \\ -\frac{1}{2}\lambda_T(\theta - \bar{\theta})^2 + o_p(\lambda_T) \\ \quad \text{if } d \in \left[-\frac{1}{2}, -\frac{1}{4}\right), \end{cases}$$

and using Assumption (e) it follows that

$$\tilde{\theta}_{j,T} \xrightarrow{p} \begin{cases} \theta_0 & \text{if } d \in \left(-\frac{1}{4}, 0\right], \\ \frac{2c_j^2 E(x_{j,i}^2)\theta_0 + \kappa\bar{\theta}}{2c_j^2 E(x_{j,i}^2) + \kappa} & \text{if } d = -\frac{1}{4}, \\ \bar{\theta} & \text{if } d \in \left[-\frac{1}{2}, -\frac{1}{4}\right). \end{cases}$$

Thus the shrinkage estimators remain consistent for θ_0 under the new null hypothesis (46) and converge in probability to $\bar{\theta}$ under the new alternative hypothesis (47). When $d = -1/4$, the shrinkage estimators converge to the weighted average of θ_0 and $\bar{\theta}$.

Next consider the asymptotic distribution of the proposed test statistic. For $d \in [-1/4, 0]$ it can be shown that the shrinkage estimator is $T^{1/2+d}$-consistent and asymptotically normal:

$$\begin{aligned} & T^{1/2+d}\left\{\tilde{\theta}_{j,T} - \theta_0 - \lambda_T[\nabla_{\theta\theta}Q_T(\theta) - \lambda_T]^{-1}(\hat{\theta}_{j,T} - \bar{\theta})\right\} \\ = \; & T^{1/2+d}\frac{\frac{Y'X_j}{T}\left(\frac{X_j'X_j}{T}\right)^{-1}\frac{X_j'u}{T}}{\frac{Y'X_j}{T}\left(\frac{X_j'X_j}{T}\right)^{-1}\frac{X_j'Y}{T} + \lambda_T} + \frac{T^{1/2+d}\lambda_T}{\frac{Y'X_j}{T}\left(\frac{X_j'X_j}{T}\right)^{-1}\frac{X_j'Y}{T} + \lambda_T}(\hat{\theta}_{j,T} - \theta_0) \\ \xrightarrow{d} \; & N\left(0, \frac{\sigma^2}{c_j^2 E(x_{j,i}^2)}\right). \end{aligned} \tag{48}$$

We also have

$$\begin{aligned} [\nabla_{\theta\theta}Q_{j,T}(\theta) - \lambda_T]^{-1} &= \left[-\frac{X_j'Y}{T}\left(\frac{X_j'X_j}{T}\right)^{-1}\frac{X_j'Y}{T} - \lambda_T\right]^{-1} \\ &= \begin{cases} -T^{-2d}[c_j^2 E(x_{j,i}^2)]^{-1} + o_p(T^{-2d}) & \text{if } d \in (-1/4, 0] \\ -T^{\frac{1}{2}}[c_j^2 E(x_{j,i}^2) + \kappa]^{-1} + o_p(T^{\frac{1}{2}}) & \text{if } d = 1/4 \end{cases}, \end{aligned} \tag{49}$$

$$
\begin{aligned}
\hat{\Sigma}_T &= \begin{bmatrix} \hat{\Pi}^2_{1,T}\frac{1}{T}\sum_{i=1}^T x^2_{1,i}\hat{u}^2_{1,i} & \hat{\Pi}_{1,T}\hat{\Pi}_{2,T}\frac{1}{T}\sum_{i=1}^T x_{1,i}x_{2,i}\hat{u}_{1,i}\hat{u}_{2,i} \\ \hat{\Pi}_{1,T}\hat{\Pi}_{2,T}\frac{1}{T}\sum_{i=1}^T x_{1,i}x_{2,i}\hat{u}_{1,i}\hat{u}_{2,i} & \hat{\Pi}^2_{2,T}\frac{1}{T}\sum_{i=1}^T x^2_{2,i}\hat{u}^2_{2,i} \end{bmatrix}, \\
&= \sigma^2 T^{2d}\begin{bmatrix} c_1^2 E(x^2_{1,i}) & 0 \\ 0 & c_2^2 E(x^2_{2,i}) \end{bmatrix} + o_p(T^{2d}) \qquad (50)
\end{aligned}
$$

where $\hat{\Pi}_{j,T} = (X_j' X_j)^{-1}X_j' Y$ and $\hat{u}_{j,t} = y_i - \hat{\theta}_{j,T}x_i$. Combining (48)–(50) we can show that

$$
\hat{R}_T \overset{d}{\to} \chi^2(1) \qquad (51)
$$

for $d \in (-1/4, 0]$. Thus, our test statistic has the same asymptotic null distribution under the more general null hypothesis (46).

Similarly, it can be shown that, if $d = -1/4$,

$$
\hat{R}_T \overset{d}{\to} K\chi^2(1) \qquad (52)
$$

where

$$
K = \frac{\frac{1}{c_1^2 E(x^2_{1,i})} + \frac{1}{c_2^2 E(x^2_{2,i})}}{\frac{c_1^2 E(x^2_{1,i})}{(c_1^2 E(x^2_{1,i})+\kappa)^2} + \frac{c_2^2 E(x^2_{2,i})}{(c_2^2 E(x^2_{2,i})+\kappa)^2}} > 1
$$

and that, if $d \in (-1/2, -1/4)$, $\hat{R}_T$ diverges at rate T^{-1-4d}.

Remarks.

1. Comparing (51) and (52), we interpret the case with $d = -1/4$ as a local alternative. Because the null distribution (51) is bounded above by the distribution under the local alternative (52), our test has nontrivial local power against the local alternative. The asymptotic local power of our test depends on two factors. First, the local power is increasing in the degree of shrinkage, κ. (52) also shows that the test would not have any local power if it is constructed from non-shrinkage extremum estimators. Second, the asymptotic local power approaches one as the strength of instruments, c_1 and c_2, approaches zero. It is interesting to note that the asymptotic local power does not depend on the choice of $\bar{\theta}$. This is because the shrinkage estimators are centered at their means.

2. The number $d = -1/4$ turns out to be special when it comes to identification. In a recent paper Antoine and Renault (2009) show that the standard Wald test is valid when the quality of instruments is mixed. They show that the fastest rate at which Π_T converges to zero must be slower than $d = -1/4$ in our notation. Thus, their conditions for the validity of Wald tests and our null hypothesis (46) coincide. Because $d = -1/4$ is exactly on the boundary in their paper and in the above analysis, we interpret the case $d = -1/4$ as a local alternative.

3. The above consistency result includes Theorem 2(b) as a special case with $d = -1/2$ and shows that our test is consistent for more general fixed alternatives than the one considered in Section 2.

4 Empirical implementation of our proposed test

The test that we propose is easy to implement even in highly-dimensional models and has the advantage of having power against weak identification. However, in order to implement the test, one needs to choose the shrinkage parameter, λ_T, while ensuring that it satisfies Assumption (e): $\lambda_T = \kappa T^{-\frac{1}{2}}$. Intuitively, the choice of κ involves a trade-off between bias and variance of the shrinkage estimator: the larger κ is, the more the shrinkage estimator is coaxed towards the pseudo parameter value $\bar{\theta}$. Thus the shrinkage parameter will be more biased and have a smaller variance for a larger value of κ. Given this bias-variance trade-off, we propose to choose κ from a finite sequence of positive numbers by a cross-validation procedure that minimizes the mean-squared error of the shrinkage estimator.[7] Note that any $\lambda_T = \hat{\kappa}_T T^{-1/2}$ such that

- Under the null hypothesis $\hat{\kappa}_T \xrightarrow{p} \kappa^*$ where κ^* is a positive constant; and
- Under the alternative hypothesis $\hat{\kappa}_T = O_p(1)$

satisfies Assumption (e) and is asymptotically valid. Therefore, we expect that the proposed estimator $\hat{\kappa}_T$ (and effectively κ^*) obtained via cross-validation will converge to the value of κ that minimizes the mean-squared error of the shrinkage estimator under the null hypothesis. Under the alternative hypothesis, $\hat{\kappa}_T$ is nonzero and is finite by construction. Our choice of κ that minimizes the mean-squared error of the shrinkage estimator may not be optimal for testing our hypothesis, but is asymptotically valid. We outline our cross-validation procedure below and investigate its small sample properties in the next section. It is left for future research to theoretically investigate the effect of cross validation on the performance of the proposed test.[8]

[7]See Stone (1974) for a theoretical analysis of cross-validation and Carrasco (2008, Section 4) for a recent application of cross-validation methods.

[8]It is possible that choosing the bandwidth by cross validation may improve the performance of the test. As an anonymous referee suggests, it could be possible that when the identification is strong, the variance of the estimator may be small even without shrinkage so that the cross validation would choose a small κ to reduce the bias. This property might make the size distortion small. On the other hand, when the identification is weak, the variance of the estimator may be large and the cross validation might choose a large κ to reduce the variance at the cost of bias

Suppose that we estimate parameters by GMM in which moment functions are serially uncorrelated when they are evaluated at the true parameter values, as in the models considered in Sections 4 and 5.

Step 0. Estimate θ by GMM:

$$\begin{aligned}\hat{\theta}_{1T} &= \arg\max_{\theta\in\Theta} Q_{1T}(\theta), \\ \hat{\theta}_{2T} &= \arg\max_{\theta\in\Theta} Q_{2T}(\theta),\end{aligned}$$

where $Q_{1T}(\theta) = \bar{m}_T(\theta)' W_{1T} \bar{m}_T(\theta)$, $Q_{2T}(\theta) = \bar{m}_T(\theta)' W_{2T} \bar{m}_T(\theta)$, $\bar{m}_T(\theta) = (1/T)\sum_{t=1}^{T} m(z_t,\theta)$, and $m(z_t,\theta)$ is a moment function satisfying $E[m(z_t,\theta_0)] = 0$ for some $\theta_0 \in \Theta$ (for example, $\hat{\theta}_{1T}$ and $\hat{\theta}_{2T}$ can be GMM estimators with identity and optimal weighting matrices).

Step 1. Pick an arbitrary value of λ_T such that $\lambda_T \in \{\lambda_{1,T}, \lambda_{2,T}, ..., \lambda_{L,T}\}$, where $\lambda_{j,T} = c_j T^{-1/2}$ for $j = 1, 2, ..L$, c_j is a positive constant, and L is finite.

Step 2. Pick an arbitrary $t \in \{1, 2, ..., T\}$.

Step 3. Use all the sample observations except t to estimate their shrinkage versions,

$$\begin{aligned}\tilde{\theta}_{1T,t} &= \arg\max_{\theta\in\Theta}\left[Q_{1T,t}(\theta) - \frac{\lambda_T}{2}\|\theta - \bar{\theta}\|^2\right], \\ \tilde{\theta}_{2T,t} &= \arg\max_{\theta\in\Theta}\left[Q_{2T,t}(\theta) - \frac{\lambda_T}{2}\|\theta - \bar{\theta}\|^2\right],\end{aligned}$$

where $Q_{1T,t}(\theta) = \bar{m}_{T,t}(\theta)' W_{1T} \bar{m}_{T,t}(\theta)$, $Q_{2T,t}(\theta) = \bar{m}_{T,t}(\theta)' W_{2T} \bar{m}_{T,t}(\theta)$ and $\bar{m}_{T,t}(\theta) = (1/(T-1))\sum_{s\neq t} m(z_s,\theta)$.

Step 4. Repeat Step 3 for $t = 1, 2, ..., T$ and construct a criterion function based on a Mean Squared Error (MSE) estimate of these parameter estimates such as[9]

$$trace\left(MSE(\lambda_T)\right) = trace\left(\sum_{s=1}^{T}\left[\tilde{\theta}_{1T,s} - \hat{\theta}_{1T}\right]\left[\tilde{\theta}_{1T,s} - \hat{\theta}_{1T}\right]' + \left[\tilde{\theta}_{2T,s} - \hat{\theta}_{2T}\right]\left[\tilde{\theta}_{2T,s} - \hat{\theta}_{2T}\right]'\right) \tag{53}$$

Step 5. Repeat steps 2-4 for all values of λ_T, thus obtaining a vector of $L \times 1$ Mean Square Error estimates:

$$\left\{trace\left(MSE(\lambda_{1,T})\right), trace\left(MSE(\lambda_{2,T})\right), ..., trace\left(MSE(\lambda_{L,T})\right)\right\}.$$

---

inflation. This might make the test more powerful. We are grateful to the anonymous referee for this conjecture.

[9]Alternatively, one could consider the determinant (as opposed to the trace) as the criterion function. In the Monte Carlo section, we will investigate both.

Step 6. Choose λ_T^* such that $\lambda_T^* = \arg\min_{l=1,\ldots,L} trace\left(MSE(\lambda_{l,T})\right).$

Step 7. Re-estimate the shrinkage estimators evaluated at λ_T^*:

$$\begin{aligned}\tilde{\theta}_{1T} &= \text{argmax}_{\theta\in\Theta}\left[Q_{1T}(\theta) - \frac{\lambda_T^*}{2}\|\theta-\bar{\theta}\|^2\right],\\ \tilde{\theta}_{2T} &= \text{argmax}_{\theta\in\Theta}\left[Q_{2T}(\theta) - \frac{\lambda_T^*}{2}\|\theta-\bar{\theta}\|^2\right],\end{aligned}$$

and evaluate the test statistic by

$$\hat{R}_T = \hat{d}_T'(\hat{D}_T'\hat{\Sigma}_T(\hat{\theta}_T)\hat{D}_T)^{-1}\hat{d}_T,$$

where

$$\begin{aligned}\hat{d}_T &= T^{\frac{1}{2}}(\tilde{\theta}_{2T} - \tilde{\theta}_{1T} - \lambda_T^*\hat{B}_{2T} + \lambda_T^*\hat{B}_{1T}),\\ \hat{D}_T' &= \begin{bmatrix} -[\nabla_{\theta\theta}Q_{1T}(\tilde{\theta}_{1T}) - \lambda_T^* I_k]^{-1} \\ [\nabla_{\theta\theta}Q_{2T}(\tilde{\theta}_{2T}) - \lambda_T^* I_k]^{-1}\end{bmatrix},\\ \hat{\Sigma}_T &= \begin{bmatrix}\hat{\Sigma}_{11,T} & \hat{\Sigma}_{12,T}\\ \hat{\Sigma}_{21,T} & \hat{\Sigma}_{22,T}\end{bmatrix},\end{aligned}$$

and $\hat{B}_{jT} = [\nabla_{\theta\theta}Q_{jT}(\tilde{\theta}_{j,T}) - \lambda_T^* I_k]^{-1}(\hat{\theta}_{jT} - \bar{\theta})$ for $j \in \{1,2\}$, $\hat{\Sigma}_T$ is a consistent estimator of the asymptotic covariance matrix of $T^{1/2}[\nabla_\theta Q_{1T}(\theta_0)'\ \nabla_\theta Q_{2T}(\theta_0)']$.

Step 8. Reject the null hypothesis of strong identification in favor of weak or no identification at significance level α if $\hat{R}_T$ is bigger than the $(1-\alpha) - th$ percentile of a χ_k^2 distribution.

5 Monte Carlo Experiments

We analyze the finite sample performance of our proposed test in two setups: the Consumption Capital Asset Pricing Model (CCAPM) and the Taylor rule monetary policy model. We will compare the performance of our test with that of Wright (2003) and discuss a cross-validation method to estimate λ_T.[10]

5.1 Consumption Capital Asset Pricing Models

In this sub-section, we investigate the finite-sample performance of the proposed test using the Consumption Capital Asset Pricing Model used in Wright (2003). Consumption and dividend

[10] For computational reasons we will not consider the method proposed by Wright (2002). In this section we let $\overline{\theta} = 0$. Unreported Monte Carlo experiments show that the procedure is robust to different choices for $\overline{\theta}$.

growth are assumed to follow a first-order Gaussian vector autoregression

$$\begin{bmatrix} \log\left(\frac{C_t}{C_{t-1}}\right) \\ \log\left(\frac{D_t}{D_{t-1}}\right) \end{bmatrix} = \mu + \Phi \begin{bmatrix} \log\left(\frac{C_{t-1}}{C_{t-2}}\right) \\ \log\left(\frac{D_{t-1}}{D_{t-2}}\right) \end{bmatrix} + \begin{bmatrix} u_{ct} \\ u_{dt} \end{bmatrix}, \tag{54}$$

where C_t is consumption, D_t is dividend, μ is a 2×1 vector, Φ is a 2×2 matrix of constants, $[u_{ct}, u_{dt}]' \overset{iid}{\sim} N(0, \Lambda)$, and (54) is approximated by a 16-state Markov chain. Then asset prices are generated so that they satisfy the Euler equation

$$E_t\left[\delta R_{t+1}\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} - 1\right] = 0, \tag{55}$$

where δ is discount factor, R_t is the gross stock return and γ is the coefficient of relative risk aversion. See Tauchen and Hussey (1991) for the quadrature method used to simulate data.

Following Wright (2003), we let $\theta \in \Theta$, where $\theta = [\delta, \gamma]'$, and $\Theta = [0.7, 1.3] \times [0, 30]$. In our notation the objective function can be written as

$$Q_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left[\delta R_{t+1}\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} - 1\right] Z_t' W_T \frac{1}{T}\sum_{t=1}^{T} Z_t \left[\delta R_{t+1}\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} - 1\right], \tag{56}$$

$Z_t = [1, R_t, C_t/C_{t-1}]'$, and W_T is a weighting matrix. We use the identity matrix for $\hat{\theta}_{1,T}$ and $\tilde{\theta}_{1,T}$ and the optimal weighting matrix for $\hat{\theta}_{2,T}$ and $\tilde{\theta}_{2,T}$.

We consider one model where the parameters are strongly identified, two models where the parameters are not (or only partially) identified, and two models where the parameters are weakly identified. See Table 1 for the parameter values in each of the five models. Model SI is a slight modification of experiment $1B$ of Tauchen (1986) and model FR of Wright (2003), in which correlation is introduced among the instruments to satisfy our assumptions (g) and (h). In model SI, the parameters are strongly identified. Models $PI1$ and $PI2$ are the same as models $RF1$ and $RF2$ of Wright (2003). In these models, the instruments C_{t+1}/C_t and D_{t+1}/D_t are independent of C_{t+1}/C_t and R_{t+1}, and the rank of the Jacobian matrix is 1. Models $WI1$ and $WI2$ are modifications of models $NRF1$ and $NRF2$ of Wright (2003) which is based on Kocherlakota (1990). In these models, Φ is the same as the one in Wright (2003) when the sample size is 90 for which the value of Φ is obtained in Kocherlakota (1990). As the sample size grows, Φ converges to the matrix of zeros, which means that the instruments become weak.

We consider three sample sizes, $T = 50, 100, 200$, and set the number of Monte Carlo replications to 1000. We select λ_T via the cross validation method discussed in Section 3. We set the set of κ in Assumption (e) to $\kappa \in \{1, 5, 10\}$ in this Monte Carlo experiment. Unlike simple parametric hypotheses, the distinction between our null and alternative hypotheses is murky in small samples. We report the median of the absolute value of the bias as well as the coverage probability of 95% confidence intervals based on t tests to assess the quality of the conventional asymptotic approximation. When identification is weak, the standard asymptotic approximation will perform poorly and we expect to see large biases and poor coverage probabilities. We compute rejection probabilities of both Wright's (2003) test as well as our $\hat{R}_T$ test at the 5% significance level. We expect Wright's (2003) test to reject the null in model SI whereas our test is expected to reject the null in models PI1, PI2, WI1 and WI2.

Table 2 shows the bias of the GMM estimators, coverage probabilities of 95% Wald confidence intervals and the rejections frequencies of Wright's (2003) test and our test implemented with a nominal size equal to 5%. As expected, the GMM estimates are highly biased and the coverage probabilities are not accurate when the parameters are not identified or weakly identified. When the parameters are strongly identified (model SI), the rejection frequencies of Wright's (2003) test increase as the sample size grows. Our proposed test is conservative in that the actual size is smaller than the nominal size. When the parameters are not identified (models $PI1$ and $PI2$), Wright's (2003) test is also conservative. Our test is powerful in that it rejects the null with probability higher than 90% even for the sample size 50. Our test has power even when the parameters are weakly identified. When the parameters are weakly identified, Wright's (2003) test rejects the null hypothesis of lack of identification 26.6%-46.6% of the times, which could mislead practitioners to believe that the model is strongly identified. While the size and power of our test does depend on the choice of κ, our test performs well when κ is chosen to minimize either the determinant or trace of MSE.

5.2 The Taylor Rule Model

We now consider the performance of a simple Taylor-rule model for monetary policy in a second series of Monte Carlo experiments. We focus on the same model that will be considered in the empirical application in the next section. The model is a simplified version of the monetary policy reaction function considered by Clarida, Gali and Gertler (2000, hereafter CGG) and it is based on the following moment conditions:[11]

$$E_t\left[\{r_t - [rr^* - (\beta - 1)\,\pi^* + \beta\pi_{t+1} + \gamma y_{t+1}]\}\,X_t\right] = 0 \qquad (57)$$

[11]The simplification consists of not considering serial correlation in the Fed Fund Rate.

where r_t is the Fed Fund Rate, π_{t+1} is the inflation rate, y_{t+1} is the average output gap between time t and $t+1$ and X_t is a vector of four instruments. We generate the instruments as: $X_t \sim N_{4\times 1}(0_{4\times 1}, \Omega_X)$ and $\Omega_X = S_X S_X'$ where S_X was set to

$$S_X = \begin{bmatrix} 1.6167 & -1.4234 & 0.1957 & -0.2524 \\ 0 & 1.9835 & -0.1077 & 0.4627 \\ 0 & 0 & 0.1635 & -0.4427 \\ 0 & 0 & 0 & 0.6272 \end{bmatrix}.$$

which had been randomly drawn.

We generate the data as follows:

$$r_t = rr^* - (\beta - 1)\pi^* + \beta\pi_{t+1} + \gamma y_{t+1} + \varepsilon_t,$$

where ε_t is $N(0,1)$, $\beta = 2$, $\gamma = 3$, rr^* is the sample average of the simulated values of $r_t - \pi_{t+1}$ (on average, it is equal to unity), and π^* is chosen such that $\overline{rr}^* \equiv rr^* - (\beta - 1)\pi^* = 1$ (which means that on average the Central Bank aims at zero inflation). The vector of regressors consists of a constant as well as $Y_t = \{\pi_{t+1}, y_{t+1}\}'$, where the latter are generated by:

$$Y_t = B_{xz} X_t + u_{X,t}$$

where $B_{xz} \equiv \vartheta\left[I_{2\times 2} \;\; 0_{2\times 2}\right]$, and $u_{X,t} \sim N_{2\times 1}(0_{2\times 1}, I_{2\times 2})$. We consider three cases: $\vartheta = 0$ (no identification, labeled "NI"), $\vartheta = T^{-1/2}$ (weak identification, labeled "WI"), $\vartheta = 1$ (strong identification, labeled "SI").

We will compare the performance of our method with that proposed by Wright (2003). In applying Wright's (2003) method, we excluded the derivative of the moment condition with respect to the constant.[12] In the no identification case, we implemented Wright's (2003) method by testing the null hypothesis that the rank is 3 against the alternative that the rank is full (equal to four). Our method was implemented with a cross-validation choice of $\lambda_T = \kappa T^{1/2}$ for values of κ within a grid from 0.1 to 100, as well as with a fixed choice for $\lambda_T = \kappa T^{1/2}$, where $\kappa = 1, 5, 10$. For the cross validation, we consider both the trace, as in (53), as well as a determinant.

Table 3 reports the results. The main findings of the previous sub-section do carry over to this case. In particular, we note that Wright's (2003) test has a tendency to reject the null hypothesis of no identification when the parameters are weakly identified. Our test, implemented with the

[12] This is necessary because the test statistic is based on the demeaned gradient of the moment conditions, and if one of the derivatives of the moment conditions is constant – which will happen if one of the instruments is a constant and one of the derivatives is constant – then the gradient will have a column of zeros.

cross-validation choice for λ_T, performs really well in terms of both size and power in small samples. Wright's (2003) method also performs well in terms of size. However, in the weak identification case, Wright's (2003) test rejects the null hypothesis of lack of strong identification 20-30% of the times, thus incorrectly concluding that the model is identified in 20-30% of the cases. In the same situation, our test, instead, does reject the null hypothesis of strong identification 50-60% of the times, thus showing quite good power properties. Finally, the cross-validation procedure significantly improves the size properties of our test in finite samples relative to the case in which λ_T is pre-determined. The only notable difference with the results in the previous sub-section is that the cross-validation implemented with the determinant (rather than the trace) sometimes improves the power of the test.

6 Is the U.S. monetary policy rule identified? An analysis of identification of the U.S. forward-looking Taylor rule.

The issue of whether the parameters of structural macroeconomic models are well identified has recently received a lot of attention. In their review, An and Schorfheide (2007) acknowledge that identification problems in DSGE models are an important issue. They note that it is difficult to directly detect identification problems in large DSGE models since the mapping from the vector of structural parameters to the reduced form parameters is highly non-linear and, typically, has to be evaluated numerically. Lack of identification, therefore, constitutes a challenge for researchers because it is unclear which features of the posterior distribution are generated by prior information on rather than by information from the sample via the likelihood. So far, the main diagnostic tool to judge the extent to which data provide information regarding the parameters of interest has been to compare the prior and the posterior estimates. The method we propose in this paper has the advantage of testing whether the model's parameters suffer from weak identification prior to estimation.

The lack of identification of the parameters of various DSGE models has been documented in several papers. Canova and Sala (2009) compare the informativeness of different estimators with respect to key structural parameters in selected DSGE models, whereas Iskrev (2007) considers the issue of parameter identification in the Smets and Wouters (2007) model. Ruge-Murcia (2007) instead examines the implications of weak identification on competing estimators of DSGE models.

A distinctive feature of interest in many DSGE models is the monetary policy reaction function.

We therefore focus on it for our analysis. Usually, the monetary policy reaction function is a Taylor rule – see Taylor (1993). CGG estimate the monetary policy reaction function by GMM based on the following moment conditions:

$$E_t\left[\left\{r_t-(1-\rho_1-\rho_2)\left[rr^*-(\beta-1)\,\pi^*+\beta\pi_{t+1}+\gamma y_{t+1}\right]-\rho_1 r_{t-1}-\rho_2 r_{t-2}\right\}X_t\right]=0. \tag{58}$$

The set of instruments X_t includes 4 lags of inflation, output gap, the Fed Fund Rate, interest rate spread, money growth, and inflation in commodity prices. Let $\theta=\{\rho_1,\rho_2,\beta,\gamma,\pi^*\}$. Note that π^* is not directly identifiable from (58); it is instead estimated as: $(\widehat{\overline{rr}^*}-\widehat{rr^*})/\left(1-\beta\right)$, where $\overline{rr}^*\equiv rr^*-(\beta-1)\,\pi^*$ and rr^* is the sample average of the real interest rate. The parameter β is typically interpreted as the "inflation-aversion" parameter, whereas γ is interpreted as the "output-gap reaction" parameter.

We follow CGG and use the same quarterly data spanning the period 1960:1-1996:4. In particular, we collect interest rate and inflation data from CITIBASE. The Fed Fund Rate is the average value in the first month of each quarter, expressed in annual rates (FYFF). The inflation rate is the annualized rate of change of the GDP deflator (GDPP) between two subsequent quarters. The output gap is from the Congressional Budget Office.

In CGG, the structural parameters have a one-to-one relationship with the parameters in a standard linear GMM moment condition:

$$E_t\left[\left\{r_t-\alpha_1-\alpha_2\pi_{t+1}-\alpha_3 r_{t-1}-\alpha_4 r_{t-2}-\alpha_5 y_{t+1}\right\}X_t\right]=0, \tag{59}$$

that is, $E\left[g_t\left(\alpha\right)\right]=0$, where $g_t\left(\alpha\right)\equiv\left(r_t-\alpha' Z_t\right)X_t$ for Z_t being the vector containing a constant, the one-step ahead inflation rate, the interest rate lagged one and two periods, and the one-step ahead output gap. The structural parameters estimates are recovered from the estimated GMM parameters via a non-linear mapping procedure. To estimate the GMM parameters, let $Q_T\left(\alpha\right)=-\frac{1}{2}\overline{g}_T\left(\alpha\right)' W\overline{g}_T\left(\alpha\right)$, where $\overline{g}_T\left(\alpha\right)=T^{-1}\sum_{t=1}^{T}g_t\left(\alpha\right)=T^{-1}\sum_{t=1}^{T}X_t r_t-T^{-1}\sum_{t=1}^{T}X_t Z_t'\alpha$, $G=T^{-1}\sum_{t=1}^{T}\partial g_t\left(\alpha\right)/\partial\alpha'=-T^{-1}\sum_{t=1}^{T}X_t Z_t'$, and $\triangledown_{\theta\theta}Q_T\left(\theta,W\right)=-G'WG$.

The shrinkage GMM estimator satisfies:

$$\begin{aligned}\widetilde{\alpha}\left(W\right) &= \arg\max_{\alpha}\left(Q_T\left(\alpha\right)-0.5\lambda||\alpha||^2\right) \\ &= \arg\max_{\alpha}\left(-\frac{1}{2}\overline{g}_T\left(\alpha\right)' W\overline{g}_T\left(\alpha\right)-0.5\lambda_T\sum_{s=1}^{5}{\alpha_s}^2\right)\end{aligned} \tag{60}$$

From (60), the first order conditions give:

$$\widetilde{\alpha}(W) = \left(G'WG + \lambda_T I_p\right)^{-1} \left(G'W \frac{1}{T} \sum_{t=1}^{T} X_t r_t\right)$$

We will consider two shrinkage estimators: $\widetilde{\alpha}_1 = \widetilde{\alpha}(W^*)$, where W^* is the inverse of the asymptotic variance of $g_t(\alpha)$, and $\widetilde{\alpha}_2 = \widetilde{\alpha}(I)$. In the implementation, we chose λ_T by using the cross validation method described in Section 3.

Panel A in Table 4 shows the empirical results for the GMM parameters, α. Our results show that we do not reject the null hypothesis of identification in both the Volker-Greenspan period as well as in the Pre-Volker period. Panel B shows instead the results for the structural parameters, θ. The results for the latter are very different, and show that we cannot reject the null of strong identification in the Pre-Volker period but we do reject identification in the Volker-Greenspan era. Our results suggest that, while identification issues are not a concern for the GMM parameters, they are indeed a concern for the structural parameters in the monetary policy reaction function. In passing, note that Mavroeidis (2010) estimates the joint confidence sets for the inflation-aversion and output gap reaction parameters by using Stock and Wright's (2000) identification-robust test.[13] His objective is rather different from ours. While we want to test whether the parameters are weakly identified, he instead wants to estimate a confidence set that is robust to weak identification.

7 Conclusions

This paper provides a new test for identification. The test has a limiting chi-square distribution under the null hypothesis of identification. Among the advantages of our test, we have: (i) the test is simple to implement; (ii) the test has power against weak identification; (iii) unlike most of the tests available in the literature, our test directly focuses on the null hypothesis of interest (identification) rather than the opposite (no identification).

We document the good small sample size and power properties of our test via Monte Carlo simulations calibrated on both a Consumption Capital Asset Pricing Model and a Taylor rule monetary policy reaction function. Finally, we implement our test to analyze whether the structural

[13] He finds that the confidence sets are much wider in the Volker-Greenspan's subsample than in the Pre-Volker era, and that the confidence sets contain parameters included in both the determinate and the indeterminacy regions, which is consistent with our results. However, his analysis is computationally very demanding, and very difficult to implement in highly dimensional parameter spaces.

parameters of the Taylor rule monetary policy reaction function are identified in the data. We show that identification is a concern mainly in the Volker-Greenspan era.

In this paper we used the quadratic penalty term. Recently Caner (2009) developed GMM estimators with least absolute shrinkage and selection operator (LASSO) under strong identification. Extending our results to non-quadratic penalty terms, such as LASSO, is an interesting avenue of research but is beyond the scope of this paper.

Appendix A: Proofs of the Theorems

Proof of Theorem 1.

Part (a): Equation (11) trivially follows from Theorem 3.1 of Newey and McFadden (1994, p.2143) and Assumptions (a), (b) and (c). Because $\lambda_T = o(1)$, it follows from Theorem 2.1 of Newey and McFadden (1994, p.2121) and Assumptions (a), (b) and (c) that $\tilde{\theta}_T = \theta_0 + o_p(1)$. The first-order condition for $\tilde{\theta}_T$ is

$$\nabla_\theta Q_T(\tilde{\theta}_T) - \lambda_T(\tilde{\theta}_T - \bar{\theta}) \;=\; 0_{k\times 1}, \tag{61}$$

By applying the mean value theorem to (61) we obtain

$$\tilde{\theta}_T - \theta_0 - \lambda_T[\nabla_{\theta\theta} Q_T(\bar{\theta}_T) - \lambda_T I_k]^{-1}\left(\theta_0 - \bar{\theta}\right) = \; -[\nabla_{\theta\theta} Q_T(\bar{\theta}_T) - \lambda_T I_k]^{-1}\nabla_\theta Q_T(\theta_0), \tag{62}$$

where $\bar{\theta}_T$ is a point between θ_0 and $\tilde{\theta}_T$. Because $Q_T(\theta)$ is twice continuously differentiable by Assumption (b), its third derivatives are bounded on the compact set Θ. Because $\tilde{\theta}_T \xrightarrow{p} \theta_0$ and $Q(\theta_0)$ is non-singular by Assumption (c.iii), $[\nabla_{\theta\theta} Q_T(\theta_0) - \lambda_T I_k]^{-1}$ is non-singular with probability approaching one. Thus,

$$\begin{aligned} &\frac{\partial}{\partial\theta}\mathrm{vec}\left\{[\nabla_{\theta\theta} Q_T(\theta) - \lambda_T I_k]^{-1}\right\} = -\{[\nabla_{\theta\theta} Q_T(\theta) - \lambda_T I_k]^{-1} \otimes \\ &[\nabla_{\theta\theta} Q_T(\theta) - \lambda_T I_k]^{-1}\}\frac{\partial}{\partial\theta}\mathrm{vec}[\nabla_{\theta\theta} Q_T(\theta)] \end{aligned} \tag{63}$$

is finite in a shrinking neighborhood of θ_0 with probability approaching one, and

$$[\nabla_{\theta\theta} Q_T(\bar{\theta}_T) - \lambda_T I_k]^{-1} \;=\; [\nabla_{\theta\theta} Q_T(\theta_0) - \lambda_T I_k]^{-1} + O_p(\|\tilde{\theta}_T - \theta_0\|). \tag{64}$$

It follows from (62) and (64) that

$$\begin{aligned} \tilde{\theta}_T - \theta_0 - \lambda_T B_T(\theta_0) \;&=\; -[\nabla_{\theta\theta} Q_T(\theta_0) - \lambda_T I_k]^{-1}\nabla_\theta Q_T(\theta_0) \\ &\quad + O_p(\lambda_T\|\tilde{\theta}_T - \theta_0\|). \end{aligned} \tag{65}$$

Therefore equation (12) follows from (65) and Assumptions (c.ii), (c.iii) and (f.i).

Equation (13) in Part (b): We will follow the proof of Theorem 1 of Stock and Wright (2000). First, we will show $\hat{\beta}_T = \beta_0 + O_p(T^{-1/2})$. Second, we will find a limiting representation for $\nabla Q_T(\alpha, \beta_0 + bT^{-1/2})$. Third, we will prove equation (13).

It follows from Assumption (d.i) that

$$Q_T(\theta) \;\xrightarrow{p}\; Q_\beta(\beta) \tag{66}$$

uniformly in θ. Because $Q_\beta(\beta)$ is uniquely maximized at β_0 by Assumption (d.iv), we can show that $\hat{\beta}_T \xrightarrow{p} \beta_0$ by using the standard argument. Next we will show that $\hat{\beta}_T = \beta_0 + O_p(T^{-1/2})$. The first order condition for maximizing $Q_T(\theta)$ with respect to β is

$$\nabla_\beta Q_T(\hat{\theta}_T) \ = \ 0_{k_2 \times 1}. \tag{67}$$

By applying the mean value theorem to (67) we obtain

$$\nabla_\beta Q_T(\theta_0) + \nabla_{\beta\alpha} Q_T(\bar{\theta}_T)(\hat{\alpha}_T - \alpha_0) + \nabla_{\beta\beta} Q_T(\bar{\theta}_T)(\hat{\beta}_T - \beta_0) \ = \ 0_{k_2 \times 1}, \tag{68}$$

where $\bar{\theta}_T = [\bar{\alpha}_T', \bar{\beta}_T']'$ is a point between θ_0 and $\hat{\theta}_T$. Because $\hat{\beta}_T \xrightarrow{p} \beta_0$, Θ is compact by Assumption (a), which implies that $\hat{\alpha}_T - \alpha_0 = O_p(1)$, $\nabla_{\beta\alpha} Q_T(\theta) = O_p(T^{-1/2})$ uniformly in θ by Assumption (d.iii), $\nabla_{\beta\beta} Q_T(\theta) - \nabla_{\beta\beta} Q_\beta(\beta) = O_p(T^{-1/2})$ uniformly in θ by Assumption (d.iv), $\nabla_{\beta\beta} Q_\beta(\beta)$ is bounded and non-singular by Assumptions (a), (b) and (c.iii) we have

$$\hat{\beta}_T - \beta_0 \ = \ O_p(T^{-1/2}). \tag{69}$$

Next we will find a limiting representation for $\nabla Q_T(\alpha, \beta_0 + bT^{-1/2})$ as an empirical process in $[a', b']' \in \Theta_A \times \overline{\Theta}_B$ where $\overline{\Theta}_B$ is a compact set in $\Re^{k_2}$. We have

$$\begin{aligned} & TQ_T(a, \beta_0 + bT^{-1/2}) \\ = \ & TQ_T(a, \beta_0) + T^{\frac{1}{2}} \nabla_\beta Q_T(a, \beta_0) b + \frac{1}{2T} b' \nabla_{\beta\beta} Q_T(a, \beta_0) b + o_p(T^{-1}) \\ \Rightarrow \ & Q_\alpha(a, \beta_0) + Z_\beta(\alpha, \beta_0)' b + \frac{1}{2} b' \nabla_{\beta\beta} Q_\beta(a, \beta_0) b. \end{aligned} \tag{70}$$

Thus by Lemma 3.2.1 of van der Vaart and Wellner (1996, p.286), we conclude that $[\hat{\alpha}_T', T^{1/2}(\hat{\beta}_T - \beta_0)']' \Rightarrow [\alpha^{*\prime}, b^{*\prime}(\alpha^*)]'$ where α^* maximizes (3) and $b^*(\alpha)$ is given in (4).

Equation (14) in Part (b): First, we will show the consistency and convergence rates of $\tilde{\alpha}_T$ and $\tilde{\beta}_T$. Because $\sup_{\theta \in \Theta} |Q_T(\theta) - Q_\beta(\beta)| = o_p(1)$, $Q_\beta(\beta)$ is uniquely maximized at β_0, and $\lambda_T = o(1)$ by Assumptions (d.iv) and (e), one can show that $\tilde{\beta}_T \xrightarrow{p} \beta_0$ by using a standard argument. It follows from the first order condition for $\tilde{\alpha}_T$,

$$\nabla_\alpha Q_T(\tilde{\alpha}_T, \tilde{\beta}_T) - \lambda_T(\tilde{\alpha}_T - \bar{\alpha}) \ = \ 0_{k_1 \times 1}, \tag{71}$$

and Assumption (d.ii) that $\tilde{\alpha}_T - \bar{\alpha} = O_p(1/(\lambda_T T^{1/2})) = O_p(1)$. An application of the mean value theorem to the first order condition for $\tilde{\beta}_T$,

$$\nabla_\beta Q_T(\tilde{\alpha}_T, \tilde{\beta}_T) - \lambda_T(\tilde{\beta}_T - \bar{\beta}) \ = \ 0_{k_1 \times 1}, \tag{72}$$

around β_0 yields

$$\begin{aligned} & \tilde{\beta}_T - \beta_0 - \lambda_T [\nabla_{\beta\beta} Q_T(\tilde{\alpha}_T, \bar{\beta}_T) - \lambda_T I_{k_2}]^{-1} (\beta_0 - \bar{\beta}) \\ = & -[\nabla_{\beta\beta} Q_T(\tilde{\alpha}_T, \bar{\beta}_T) - \lambda_T I_{k_2}]^{-1} \nabla_\beta Q_T(\tilde{\alpha}_T, \beta_0), \end{aligned} \tag{73}$$

where $\bar{\beta}_T$ is a point between $\tilde{\beta}_T$ and β_0. By Assumptions (d.i), (d.iv) and (e), (73) can be written as

$$\begin{aligned} & \tilde{\beta}_T - \beta_0 - \lambda_T [\nabla_{\beta\beta} Q_\beta(\beta_0)]^{-1} (\beta_0 - \bar{\beta}) \\ = & -[\nabla_{\beta\beta} Q_\beta(\beta_0)]^{-1} \nabla_\beta Q_T(\bar{\alpha}, \beta_0) + O_p(T^{-\frac{1}{2}} \|\tilde{\alpha}_T - \overline{\alpha}\|) + O_p(T^{-\frac{1}{2}}). \end{aligned} \tag{74}$$

It follows from (74) and Assumptions (a), (d.i), (d.iv) and (e) that

$$\tilde{\beta}_T - \beta_0 = O_p(T^{-1/2}). \tag{75}$$

It follows from (71), (75) and Assumptions (d.ii) and (d.iii) that

$$\begin{aligned} \tilde{\alpha}_T - \bar{\alpha} &= \frac{1}{\lambda_T} \nabla_\alpha Q_T(\tilde{\alpha}_T, \beta_0) + \frac{1}{\lambda_T} \nabla_{\alpha\beta} Q_T(\tilde{\alpha}_T, \bar{\beta}_T)(\tilde{\beta}_T - \beta_0) \\ &= O_p\left(\frac{1}{\lambda_T T}\right), \end{aligned} \tag{76}$$

where $\bar{\beta}_T$ is a point between $\tilde{\beta}_T$ and β_0.

Second we will consider a limiting representation for

$$\begin{aligned} & Q_T\left(\bar{\alpha} + \frac{a}{\lambda_T T}, \beta_0 + \lambda_T [\nabla_{\beta\beta} Q_\beta(\beta_0)]^{-1} (\beta_0 - \bar{\beta}) + T^{-\frac{1}{2}} b\right) + \frac{\lambda_T}{2} \times \\ \times & \left(\left\|\frac{a}{\lambda_T T}\right\|^2 + \|\beta_0 + \lambda_T [\nabla_{\beta\beta} Q_\beta(\beta_0)]^{-1} (\beta_0 - \bar{\beta}) + T^{-\frac{1}{2}} b - \bar{\beta}\|^2\right) \end{aligned} \tag{77}$$

as an empirical process in $[a', b']' \in \overline{\Theta}_A \times \overline{\Theta}_B$ where $\overline{\Theta}_A \times \overline{\Theta}_B$ is a compact set in $\Re^{k_1} \times \Re^{k_2}$. By using Taylor's theorem, (77) can be written as

$$\begin{aligned} & Q_T(\bar{\alpha}, \beta_0) + \frac{\lambda_T}{2} \|\beta_0 - \bar{\beta}\|^2 \\ & + \left[\nabla_\theta Q_T(\bar{\alpha}, \beta_0) - \lambda_T \begin{pmatrix} 0_{k_1 \times 1} \\ \beta_0 - \bar{\beta} \end{pmatrix}\right]' \begin{bmatrix} \frac{a}{\lambda_T T} \\ \lambda_T [\nabla_{\beta\beta} Q_\beta(\beta_0)]^{-1} (\beta_0 - \bar{\beta}) + T^{-\frac{1}{2}} b \end{bmatrix} \\ & + \frac{1}{2} \begin{bmatrix} \frac{a}{\lambda_T T} \\ \lambda_T [\nabla_{\beta\beta} Q_\beta(\beta_0)]^{-1} (\beta_0 - \bar{\beta}) + T^{-\frac{1}{2}} b \end{bmatrix}' [\nabla_{\theta\theta} Q_T(\bar{\alpha}_T, \bar{\beta}_T) - \lambda_T I_k] \\ & \times \begin{bmatrix} \frac{a}{\lambda_T T} \\ \lambda_T [\nabla_{\beta\beta} Q_\beta(\beta_0)]^{-1} (\beta_0 - \bar{\beta}) + T^{-\frac{1}{2}} b \end{bmatrix}, \end{aligned} \tag{78}$$

where $[\bar{\alpha}_T' \ \bar{\beta}_T']'$ is a point between $[\bar{\alpha}' + a'/(\lambda_T T), \beta_0' + \lambda_T(\beta_0 - \bar{\beta})'[\nabla_{\beta\beta}Q_\beta(\beta_0)]^{-1} + T^{-\frac{1}{2}}b']'$ and $[\bar{\alpha}', \beta_0']'$. Thus it follows from (78) and Assumptions (d) and (e) that

$$
\begin{aligned}
& T\left[Q_T\left(\bar{\alpha} + \frac{a}{\lambda_T T}, \beta_0 + \lambda_T[\nabla_{\beta\beta}Q_\beta(\beta_0)]^{-1}(\beta_0 - \bar{\beta}) + T^{-\frac{1}{2}}b\right)\right. \\
& + \frac{\lambda_T}{2}\left(\left\|\frac{a}{\lambda_T T}\right\|^2 + \|\beta_0 + \lambda_T[\nabla_{\beta\beta}Q_\beta(\beta_0)]^{-1}(\beta_0 - \bar{\beta}) + T^{-\frac{1}{2}}b - \bar{\beta}\|^2\right) \\
& \left. - Q_T(\bar{\alpha}, \beta_0) + \frac{\lambda_T}{2}\|\beta_0 - \bar{\beta}\|^2\right] \\
\Rightarrow \ & [Z_\beta(\bar{\alpha}, \beta_0) - \kappa(\beta_0 - \bar{\beta})]'(\kappa[\nabla_{\beta\beta}Q_\beta(\beta_0)]^{-1}(\beta_0 - \bar{\beta}) + b) \\
& + \frac{1}{2}(\kappa[\nabla_{\beta\beta}Q_\beta(\beta_0)]^{-1}(\beta_0 - \bar{\beta}) + b)'\nabla_{\beta\beta}Q_\beta(\beta_0)(\kappa[\nabla_{\beta\beta}Q_\beta(\beta_0)]^{-1}(\beta_0 - \bar{\beta}) + b).
\end{aligned}
\tag{79}
$$

Therefore Theorem 1(b) follows from Lemma 3.2.1 of van der Vaart and Wellner (1996, p.286), (76), (79) and Assumption (d.ii).

Proof of Theorem 2.

Part (a): It follows from Assumptions (c.ii), (c.iii) and (e), Theorem 1(a) and equation (65) that

$$
\begin{aligned}
& T^{\frac{1}{2}}[\tilde{\theta}_{2T} - \tilde{\theta}_{1T} - \lambda_T\hat{B}_{2T} + \lambda_T\hat{B}_{1T}] \\
= \ & -T^{\frac{1}{2}}[\nabla_{\theta\theta}Q_{2T}(\theta_0) - \lambda_T I_k]^{-1}\nabla_\theta Q_{2T}(\theta_0) + T^{\frac{1}{2}}[\nabla_{\theta\theta}Q_{1T}(\theta_0) - \lambda_T I_k]^{-1}\nabla_\theta Q_{1T}(\theta_0) \\
& -\lambda_T T^{\frac{1}{2}}(\hat{B}_{2T} - B_T) + \lambda_T T^{\frac{1}{2}}(\hat{B}_{1T} - B_T) + O_p(T^{-\frac{1}{2}}) \\
\stackrel{d}{\to} \ & N(0_{k\times 1}, D'\Sigma D).
\end{aligned}
\tag{80}
$$

Since $\hat{D}_T \stackrel{p}{\to} D$ and $\hat{\Sigma}_T \stackrel{p}{\to} \Sigma$ by Assumptions (c.iii), (e) and (g) and Theorem 1(a), the desired result follows from (80).

Part (b): First we will show a result which will be used in the subsequent proofs. Using equations (6) and (7) of Magnus and Neudecker (1999, p.11), result 0.7.4 of Horn and Johnson (1985, p.19) and Assumption (d), we obtain

$$
\begin{aligned}
& [\nabla_{\theta\theta}Q_{jT}(\theta) - \lambda_T I_k]^{-1} \\
= \ & \begin{bmatrix} T^{-1}H_{j,\alpha\alpha}(\theta) - \lambda_T I_{k_1} + o_p(T^{-1}) & T^{-1/2}H_{j,\alpha\beta}(\theta) + o_p(T^{-1/2}) \\ T^{-1/2}H_{j,\beta\alpha}(\theta) + o_p(T^{-1/2}) & \nabla_{\beta\beta}Q_{j,\beta}(\beta) - \lambda_T I_{k_2} + O_p(T^{-1/2}) \end{bmatrix}^{-1} \\
= \ & \begin{bmatrix} -\frac{1}{\lambda_T}I_{k_1} + O_p\left(\frac{1}{\lambda_T^2 T}\right) & O_p\left(\frac{1}{\lambda_T T^{1/2}}\right) \\ O_p\left(\frac{1}{\lambda_T T^{1/2}}\right) & [\nabla_{\beta\beta}Q_{j,\beta}(\beta) - \lambda_T I_{k_2}]^{-1} + O_p(T^{-\frac{1}{2}}) \end{bmatrix},
\end{aligned}
\tag{81}
$$

and

$$\lambda_T \hat{B}_{jT} \;=\; \left[\begin{array}{c} -\alpha_j^* + \bar{\alpha} + O_p(\frac{1}{\lambda_T T}) \\ \lambda_T [\nabla_{\beta\beta} Q_{j,\beta}(\beta_0) - \lambda_T I_{k_2}]^{-1}(\beta_0 - \bar{\beta}) + O_p(\frac{\lambda_T}{T^{1/2}}) \end{array} \right]. \tag{82}$$

It follows from Theorem 1(b) and equations (81) and (82) that

$$T^{-\frac{1}{2}} \hat{d}_T \;=\; \tilde{\theta}_{2,T} - \tilde{\theta}_{1,T} - \lambda_T \hat{B}_{2,T} + \lambda_T \hat{B}_{1,T} \;\Rightarrow\; \left[\begin{array}{c} \alpha_2^* - \alpha_1^* \\ 0_{k_2 \times 1} \end{array} \right] \tag{83}$$

and that

$$\begin{aligned} & \left[I_2 \otimes \left(\begin{array}{cc} T^{-\frac{1}{2}} I_{k_1} & 0_{k_1 \times k_2} \\ 0_{k_2 \times k_1} & I_{k_2} \end{array} \right) \right] \hat{D}_T \\ \xrightarrow{d} \; & \left(\begin{array}{c} -\left(\begin{array}{cc} -\frac{1}{\kappa} I_{k_1} & 0_{k_1 \times k_2} \\ \frac{1}{\kappa} (\nabla_{\beta\beta} Q_{1,\beta}(\beta_0))^{-1} H_{1,\beta\alpha}(\bar{\alpha}, \beta_0) & (\nabla_{\beta\beta} Q_{1,\beta}(\beta_0))^{-1} \end{array} \right) \\ \left(\begin{array}{cc} -\frac{1}{\kappa} I_{k_1} & 0_{k_1 \times k_2} \\ \frac{1}{\kappa} (\nabla_{\beta\beta} Q_{2,\beta}(\beta_0))^{-1} H_{2,\beta\alpha}(\bar{\alpha}, \beta_0) & (\nabla_{\beta\beta} Q_{2,\beta}(\beta_0))^{-1} \end{array} \right) \end{array} \right) \\ \equiv \; & \left[I_2 \otimes \left(\begin{array}{cc} \frac{1}{\kappa} I_{k_1} & 0_{k_1 \times k_2} \\ 0_{k_2 \times k_1} & I_{k_2} \end{array} \right) \right] \left[\begin{array}{c} -M_1 \\ M_2 \end{array} \right] \end{aligned} \tag{84}$$

By Assumption (h.ii),

$$\begin{aligned} & \left[I_2 \otimes \left(\begin{array}{cc} T^{\frac{1}{2}} I_{k_1} & 0_{k_1 \times k_2} \\ 0_{k_2 \times k_1} & I_{k_2} \end{array} \right) \right]' \hat{\Sigma}_T \left[I_2 \otimes \left(\begin{array}{cc} T^{\frac{1}{2}} I_{k_1} & 0_{k_1 \times k_2} \\ 0_{k_2 \times k_1} & I_{k_2} \end{array} \right) \right] \\ \Rightarrow \; & \left[\begin{array}{cc} \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{21}^* & \Sigma_{22}^* \end{array} \right] \end{aligned} \tag{85}$$

The result (32) follows from equations (83), (84) and (85) and equation (7) of Magnus and Neudecker (1999, p.11). The result (33) follows because the reciprocal of the eigenvalues of $\hat{D}_T' \hat{\Sigma}_T \hat{D}_T$ diverges to infinity.

References

Acemoglu, D. and S. Johnson, 2007, Disease and development: The effect of life expectancy on economic growth. Journal of Political Economy 115, 925–985.

Ait-Sahalia, Y., J.A. Parker, and M. Yogo, 2004, Luxury goods and the equity premium. Journal of Finance 59, 2959–3004.

Alcala, F. and A. Ciccone, 2004, Trade and productivity. Quarterly Journal of Economics 119. 613–646.

Amemiya, T., 1985, Advanced econometrics. Harvard University Press, Cambridge, MA.

An, S. and F. Schorfheide, 2007, Bayesian analysis of DSGE models. Econometric Reviews 26, 113–172.

Anderson, T.W., 1951, Estimating linear restrictions on regression coefficients for multivariate normal distributions. Annals of Mathematical Statistics 22, 327–351.

Anderson, T.W., 1984, An introduction to multivariate statistical analysis. Wiley, New York, NY.

Anderson, T.W. and N. Kunitomo, 1992, Tests of overidentification and predeterminedness in simultaneous equation models. Journal of Econometrics 54, 49–79.

Anderson, T.W. and N. Kunitomo, 1994, Asymptotic robustness of tests of overidentification and predeterminedness. Journal of Econometrics 62, 383–414.

Anderson, T.W. and N. Kunitomo, 2009, Likelihood ratio tests in reduced rank regression with applications to econometric structural equations. Unpublished manuscript, www-stat.stanford.edu/˜ckirby/ted/papers/ 2009_LRTsRRR.pdf.

Anderson, T.W. and H. Rubin, 1949, Estimation of the parameters of a single equation in a complete system of stochastic equations. Annals of Mathematical Statistics 20, 46–63.

Anderson, T.W. and H. Rubin, 1950, The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations. Annals of Mathematical Statistics 21, 570–582.

Antoine, B. and E. Renault, 2008, Efficient minimum distance estimation with multiple rates of convergence. Unpublished manuscript, Simon Fraser University and University of North Carolina, Chapel Hill.

Bound, J., D.A. Jaeger, and R.M. Baker, 1995, Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak. Journal of the American Statistical Association 90, 443–50.

Caner, M., 2009, LASSO type GMM estimator. Econometric Theory 25, 1–23.

Canova, F. and L. Sala, 2009, Back to square one: identification issues in DSGE models. Journal of Monetary Economics 56, 431–449

Carrasco, M., 2008, A regularization approach to the many instruments problem. Unpublished manuscript, Université de Montréal.

Clarida, R., J. Galí, and M. Gertler, 2000, Monetary policy rules and macroeconomic stability: evidence and some theory. Quarterly Journal of Economics 115, 147–180.

Consolo, A. and C.A. Favero, 2009, Monetary policy inertia: More a fiction than a fact? Journal of Monetary Economics 56: 900–906.

Cragg, J.G. and S.G. Donald, 1993, Testing identifiability and specification in instrumental variables models. Econometric Theory 9, 222–240.

Cragg, J.G. and S.G. Donald, 1996, On the asymptotic properties of LDU-based tests of the rank of a matrix. Journal of the American Statistical Association 91, 1301–1309.

Cragg, J.G. and S.G. Donald, 1997, Inferring the rank of a matrix. Journal of Econometrics 76, 223–250.

Crowe, C., 2010, Testing the transparency benefits of inflation targeting: Evidence from private sector forecasts. Journal of Monetary Economics 57, 226-232.

DeJuan, J.P. and J.J. Seater, 2007, Testing the cross-section implications of Friedman's permanent income hypothesis. Journal of Monetary Economics 54, 820–849.

Doyle, J.J., 2007, Child protection and child outcomes: Measuring the effects of foster care. American Economic Review 97, 1583–1610.

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 Dynamics and Control 30, 1707–1727.

Durbin, J., 1959, Errors in variables, Review of the International Statistical Institute 22, 23–32.

Eichenbaum, M.L., L.P. Hansen and K.J. Singleton, 1985, A time series analysis of representative agent models of consumption and leisure choice under uncertainty. Quarterly Journal of Economics 103, 51–78.

Faria, H.J. and H.M. Montesinos, 2009, Does economic freedom cause prosperity? An IV approach. Public Choice 141, 103–127.

Fuhrer, J.C. and G.D. Rudebusch, 2004, Estimating the Euler equation for output. Journal of Monetary Economics 51, 1133–1153.

Gallant, R.A. and H. White, 1988, A unified theory of estimation and inference for nonlinear dynamic models. Basil Blackwell, Oxford, UK.

Gill, L. and A. Lewbel, 1992, Testing the rank and definiteness of estimated matrices with applications to factor, state-space and ARMA models. Journal of the American Statistical Association 87, 766–776.

Guggenberger, P. and R.J. Smith, 2005, Generalized empirical likelihood estimators and tests under partial, weak and strong identification. Econometric Theory 21, 667–709.

Hahn, J. and J.A. Hausman, 2002, A new specification test for the validity of instrumental variables. Econometrica 70, 163–189.

Hahn, J., J. Ham and H.R. Moon, 2011, The Hausman test and weak instruments. Journal of Econometrics 160, 289–299.

Hansen, L.P., 1982, Large sample properties of generalized methods of moments estimators. Econometrica 50, 1029–1054.

Hausman, J.A., 1978, Specification tests in econometrics. Econometrica 46, 1251–1272.

Hayashi, F., 2000, Econometrics. Princeton University Press, Princeton, NJ.

Hoerl, A.E. and R.W. Kennard, 1970a, Ridge regression: biased estimation for non-orthogonal problems. Technometrics 12, 55–67.

Hoerl, A.E. and R.W. Kennard, 1970b, Ridge regression: applications to non-orthogonal problems. Technometrics 12, 69–82.

Horn, R.A. and C.R. Johnson, 1985, Matrix analysis. Cambridge University Press, Cambridge, UK.

Hsiao, C., 1983, Identification, in: Griliches, Z. and M.D. Intriligator, (Eds.), Handbook of econometrics, Vol. 1. North Holland, Amsterdam, The Netherlands.

Iskrev, N. 2007, How much do we learn from the estimation of DSGE models? A case study of identification issues in a New Keynesian business cycle model. Unpublished manuscript, University of Michigan.

Kleibergen, F. and R. Paap, 2006, Generalized reduced rank tests using the singular value decomposition. Journal of Econometrics 133, 97–126.

Kocherlakota, N., 1990, On tests of representative consumer asset pricing models. Journal of Monetary Economics 26, 285–304.

Koopmans, T.C. and W.C. Hood, 1953, The estimation of simultaneous linear economic relationships, in: W.C. Hood and T.C. Koopmans, (Eds.), Studies in econometric methods. Yale University Press, New Haven, CT, Chapter 6.

Krause, M.U., D. Lopez-Salido and T.A. Lubik, 2008, Inflation dynamics with search frictions: A structural econometric analysis. Journal of Monetary Economics 55, 892–916.

Magnus, J.R. and H. Neudecker, 1999, Matrix differential calculus with applications in statistics and econometrics, Revised Edition, John Wiley & Sons: Chichester, UK.

Mavroeidis, S., 2010, Monetary policy rules and macroeconomic stability: Some new evidence. American Economic Review 100, 491–503.

Nason, J.M. and G.W. Smith, 2005, Identifying the New Keynesian Phillips curve. Federal Reserve Board of Atlanta, Working Paper 2005-1.

Nelson, C.R. and R. Startz, 1990a, Some further results on the exact small sample properties of the instrumental variable estimator. Econometrica 58, 967–976.

Nelson, C.R. and R. Startz, 1990b, The distribution of the instrumental variable estimator and its t ratio when the instrument is a poor one. Journal of Business 63, S125–S140.

Newey, W.K., 1985, Generalized method of moments specification testing. Journal of Econometrics 29, 229–256.

Newey, W.K. and D.M. McFadden, 1994, Large sample estimation and hypothesis testing, in: Robert F. Engle and Daniel L. McFadden, (Eds.), Handbook of econometrics, Vol. IV. Elsevier, Amsterdam, the Netherlands.

Okui, R., 2007, Instrumental variables estimation in the presence of many moment conditions. Journal of Econometrics, forthcoming.

Park, C. and C. Kang, 2008, Does education induce healthy lifestyle? Journal of Health Economics 27, 1516–1531.

Robin, J.-M. and R. Smith, 2000, Tests of rank. Econometric Theory 16, 151–175.

Ruge-Murcia, F., 2007, Methods to estimate dynamic stochastic general equilibrium models. Journal of Economic Dynamics and Control 31, 2599–2636.

Sargan, J.D., 1958, The estimation of economic relationships using instrumental variables. Econometrica 26, 393–415.

Shapiro, A.H., 2008, Estimating the new Keynesian Phillips curve: A vertical production chain approach. Journal of Money, Credit and Banking 40, 627–666.

Smets, F. and R. Wouters, 2005, Shocks and frictions in U.S. business cycles: A Bayesian DSGE approach. American Economic Review 97, 586–606.

Staiger, D. and J.H. Stock, 1997, Instrumental variables regressions with weak instruments. Econometrica 65, 557–586.

Stock, J.H. and J.H. Wright, 2000, GMM with weak identification. Econometrica 68, 1055–1096.

Stock, J.H. and M. Yogo, 2005, Testing for weak instruments in linear IV regression, in: Donald W.K. Andrews and James H. Stock, (Eds.), Identification and inference for econometric models: Essays in honor of Thomas Rothenberg. Cambridge University Press, New York, NY, pp. 80–108.

Stone, C.J., 1974, Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society 36, 111–147.

Tauchen, G.E., 1986, Statistical properties of generalized method of moments estimators of structural parameters obtained from financial market data. Journal of Business and Economic Statistics 4, 397–425.

Tauchen, G.E. and R. Hussey, 1991, Quadrature based methods for obtaining approximate solutions to nonlinear asset pricing models. Econometrica 59, 371–396.

Taylor, J.B., 1993, Discretion versus policy rules in practice. Carnegie-Rochester Conference Series on Public Policy 39, 195–214.

Temple, J. and L. Wossmann, 2006, Dualism and cross-country growth regressions. Journal of Economic Growth 11, 187–228.

Van der Vaart, A.W., 1998, Asymptotic statistics. Cambridge University Press, Cambridge, UK.

Van der Vaart, A.W., and J.A. Wellner, 1996, Weak convergence and empirical processes with applications to statistics. Springer-Verlag, New York, NY.

Wossmann, L., and M. West, 2006, Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS. European Economic Review 50, 695-736.

Wright, J.H., 2002, Testing the null identification in GMM. Board of Governors of the Federal Reserve System, International Finance Discussion Paper Number 732.

Wright, J.H., 2003, Detecting Lack of Identification in GMM. Econometric Theory 19, 322–330.

Wu, D.-M., 1973, Alternative Tests of Independence between Stochastic Regressors and Disturbances. Econometrica 41, 733–750.

Yogo, M., 2004, Estimating the elasticity of intertemporal substitution when instruments are weak. Review of Economics and Statistics 86, 797–810.

Table 1. Parameter Values in the Models

Model	μ	Φ	Λ	δ	γ
SI	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} -0.5 & 0.1 \\ 0.1 & -0.5 \end{bmatrix}$	$\begin{bmatrix} 0.01 & 0.005 \\ 0.005 & 0.01 \end{bmatrix}$	0.97	1.3
PI1	$\begin{bmatrix} 0.018 \\ 0.013 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.0012 & 0.0017 \\ 0.0017 & 0.0146 \end{bmatrix}$	0.97	1.3
PI2	$\begin{bmatrix} 0.018 \\ 0.013 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.0012 & 0.0017 \\ 0.0017 & 0.0146 \end{bmatrix}$	1.139	13.7
WI1	$\begin{bmatrix} 0.021 \\ 0.004 \end{bmatrix}$	$\left(\frac{90}{T}\right)^{\frac{1}{2}} \begin{bmatrix} -0.161 & 0.017 \\ 0.414 & 0.117 \end{bmatrix}$	$\begin{bmatrix} 0.0012 & 0.0017 \\ 0.0017 & 0.0146 \end{bmatrix}$	0.97	1.3
WI2	$\begin{bmatrix} 0.021 \\ 0.004 \end{bmatrix}$	$\left(\frac{90}{T}\right)^{\frac{1}{2}} \begin{bmatrix} -0.161 & 0.017 \\ 0.414 & 0.117 \end{bmatrix}$	$\begin{bmatrix} 0.0012 & 0.0017 \\ 0.0017 & 0.0146 \end{bmatrix}$	1.139	13.7

Notes: μ, Φ and Λ are the intercept, matrix of slope coefficients and covariance matrix of the disturbance term, respectively, of the VAR(1) model of consumption and dividend growth.

Table 2. Rejection Frequencies of

Wright's (2003) Test and the Proposed Test

		δ		γ		Wright	Our Test				
Model	T	bias	coverage	bias	coverage	(2003)	κ= 1	κ= 5	κ= 10	det	trace
FR	50	0.007	.926	0.138	.893	.741	.070	.101	.154	.070	.070
	100	0.005	.936	0.100	.917	.913	.047	.068	.102	.047	.047
	200	0.003	.946	0.071	.936	.976	.038	.053	.081	.038	.038
RF1	50	0.034	.986	1.895	.996	.138	.899	.933	.965	.900	.900
	100	0.034	.992	1.906	.999	.113	.945	.973	.986	.945	.945
	200	0.032	1.00	1.850	.999	.097	.952	.977	.989	.952	.952
RF2	50	0.138	.415	12.200	.244	.133	.941	.967	.980	.941	.941
	100	0.138	.415	12.156	.249	.113	.988	.991	.998	.988	.988
	200	0.137	.410	12.189	.249	.100	.999	1.00	1.00	.999	.999
NRF1	50	0.012	.992	0.615	.681	.321	.558	.658	.759	.557	.558
	100	0.008	.994	0.600	.670	.310	.658	.745	.846	.658	.658
	200	0.005	.984	0.532	.686	.357	.772	.856	.923	.772	.772
NRF2	50	1.029	.126	14.585	.762	.466	.955	.977	.992	.971	.962
	100	1.139	.165	13.318	.746	.357	.918	.951	.975	.937	.926
	200	0.724	.913	3.196	.992	.266	.989	1.00	1.00	1.00	.995

Notes: The table reports median absolute biases (labeled "bias"), coverage probabilities of 95% confidence intervals (labeled "coverage"), and empirical rejection probabilities of the tests (last six columns). "Our Test" denotes our proposed $\widehat{R}_T$ test, eq. (19) ; it is either implemented with a cross-validation method for the choice of λ based on the trace, labeled "trace", or on the determinant, labeled "det". "$\kappa = 1, 5, 10$" refers to the proposed test implemented with a pre-determined choice of $\lambda = \kappa T^{-1/2}$. "Wright (2003)" is the test proposed by Wright (2003).

Table 3. Rejection frequencies of Wright's (2003) Test and the proposed test – Monetary policy example

T	Model	Wright (2003)	Our Test				
			κ=1	κ=5	κ=10	det	trace
50	SI	1	0.06	0.19	0.35	0.06	0.06
	WI	0.19	0.53	0.81	0.90	0.99	0.53
	NI	0.04	0.69	0.91	0.96	1	0.69
100	SI	1	0.06	0.14	0.24	0.06	0.06
	WI	0.26	0.58	0.86	0.93	0.99	0.58
	NI	0.04	0.74	0.94	0.97	1	0.74
200	SI	1	0.05	0.07	0.13	0.05	0.05
	WI	0.41	0.65	0.90	0.95	0.99	0.65
	NI	0.07	0.80	0.95	0.98	0.99	0.80

Notes. The table reports empirical rejection rates of nominal 5% tests for different sample sizes and for the cases of strong identification ("SI"), weak identification ("WI") and no identification ("NI"). "Our Test" denotes our proposed $\widehat{R}_T$ test, eq. (19); it is either implemented with a cross-validation method for the choice of λ based on the trace, labeled "trace", or on the determinant, labeled "det". "$\kappa = 1, 5, 10$" refers to the proposed test implemented with a pre-determined choice of $\lambda = \kappa T^{-1/2}$. "Wright (2003)" is the test proposed by Wright (2003).

Table 4. Empirical results

	Pre-Volker 1960:1-1979:2	Volker-Greenspan 1979:3-1996:4
Panel A. GMM parameters		
$\widehat{S}_T(\alpha)$	4.86	7.13
p-value	0.43	0.21
Panel B. Structural parameters		
$\widehat{S}_T(\theta)$	4.77	17.74
p-value	0.44	0.002

Notes: The table reports the value of our test statistic $\widehat{S}_T(\theta)$ and its p-values for testing identification in both the GMM parameters (Panel A) and the Structural parameters (Panel B) in the two sub-samples of interest.