Testing for Weak Identification in Possibly Nonlinear Models

Atsushi Inoue[*] Barbara Rossi[†‡]
NCSU Duke University

October 7, 2008

Abstract

In this paper we propose a chi-square test for identification. Our proposed test statistic is based on the distance between two bias-corrected shrinkage extremum estimators. The two estimators converge in probability to the same limit when identification is strong, and they converge weakly to different random variables 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 experiment results. We apply the proposed technique to test whether the structural parameters of a representative Taylor-rule monetary policy reaction function are identified.

Keywords: GMM, Shrinkage, Weak Identification

---

[*]Department of Agricultural and Resource Economics, North Carolina State Univerisity, Raleigh, NC 27695-8109. 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 Craig Burnside, Mehmet Caner and participants of the first ERID conference for comments. We thank seminar participants at the University of Washington, Seattle for helpful comments.

1 Introduction

Identifiability of a growing number of macroeconomic models has been questioned in the recent literature. See Canova and Sala (2007), Iskrev (2007) and Ruge-Murcia (2007) for empirical evidence in dynamic stochastic general equilibrium (DSGE) models, Mavroeidis (2007) 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. Identification has been a central and recurrent issue in empirical macroeconomics since Liu (1960). 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.

Since the Hessian matrix of objective functions has less than full rank when parameters are not identified, an alternative to the approach taken in this paper would be to attempt to test the null hypothesis of weak identification of nonlinear models by using tests of matrix rank, such as Cragg and Donald (1996, 1997), Gill and Lewbel (1992), Robin and Smith (2000) and Kleibergen and Paap (2006). Under the null hypothesis of weak identification in which the rank of the Hessian matrix is asymptotically less than full, the limiting distribution of extremum estimators depends on some nuisance parameters that cannot be consistently estimated, as shown by Staiger and Stock (1997) for instrumental variables (IV) estimators, and by Stock and Wright (2000) for generalized method of moments (GMM) estimators. Consequently, the limiting null distribution of a test statistic for testing the rank of a matrix will depend on the nuisance parameters. For this reason, existing matrix rank tests cannot be directly applied to test for weak identification.

Existing tests for identification deal with the nuisance parameter problem by choosing the nuisance parameters in a conservative way (Stock and Yogo, 2000), by switching the null and alternative hypotheses (Hahn and Hausman, 2002; Wright, 2002), or by restricting the nuisance parameter to take a specific value (Wright, 2003). Stock and Yogo (2000) develop tests for identification of linear instrumental variables models and select the degree of simultaneity by maximizing bias or size. Hahn and Hausman (2002) propose a test statistic by comparing forward and backward regression estimators for testing the null hypothesis of strong identification against the alternative of weak identification in linear instrumental variables models.[1] However, the Hessian of objective functions depends on nuisance parameters in nonlinear models such as the aforementioned Euler equations and DSGE models, and it is unclear whethet it is possible at all to extend the methods of Stock and Yogo (2000) and Hahn and Hausmann (2002) to nonlinear models. Our test can instead be applied to both linear and non-linear models.

The tests of Stock and Yogo (2000) and Hahn and Hausman (2002) are designed for linear IV models. To our knowledge, Wright (2002, 2003) develops the first and only formal tests for identification in nonlinear models. As pointed out by Stock, Wright and Yogo (2002, Section 7.2), however, the test of Wright (2003) is designed for testing for no and partial identification (see Choi and Phillips, 1992, for partial identification), not for testing weak identification. The test of no and partial identification is similar to testing the null hypothesis of no identification, that is, the population first-stage regression coefficient matrix is exactly zeros, by applying the conventional F test in the linear case. If the parameters are weakly identified, the size of Wright's (2003) test will be distorted in general.[2] Our test, instead, aims

[1] As pointed out by Hausman, Stock and Yogo (2005), the test statistic of Hahn and Hausman (2002) is bounded in probability under the alternative hypothesis and thus their test is not consistent.

[2] Wright (2003, Assumption 1) assumes that the Jacobian is asymptotically zero mean Gaussian. If the parameter is only weakly identified in the sense of Stock and

exactly at testing the null of identification against the alternative of weak identification.

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 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.

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 to test 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

---

Wright (2000), the Jacobian converges weakly to a Gaussian process with possibly nonzero mean, and it follows from Theorem 2 of Cragg and Donald (1997, p.228) that his statistic converges weakly to a noncentral χ^2 distribution where the noncentral parameter depends on unknown nuisance parameters. Therefore, the size of Wright's (2003) test can be distorted when parameters are weakly identified. See Dufour (1997) for this issue in a more general context.

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 $\theta_0 \in \Theta$.

Our objective is to test the null hypothesis that the parameter θ_0 is strongly identified against the alternative hypothesis that some parameters are 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 $\theta_0 = [\alpha_0', \beta_0']' \in \mathrm{int}(\Theta)$, 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.

(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 $Y = X\Pi + V$, with V being a $T \times k$ matrix of error terms.

   In their model our null and alternative hypotheses simplify to

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

   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

$$\begin{aligned} Q_T(\theta) &= -\left[\frac{1}{T}\sum_{t=1}^{T} \phi_t(\theta)\right]' \hat{W}_T \left[\frac{1}{T}\sum_{s=1}^{T} \phi_s(\theta)\right], && (7)\\ Q_\alpha(\theta) &= -m_1(\theta)' W m_1(\theta), && (8)\\ Q_{\alpha\beta}(\theta) &= -2 m_1(\theta)' W m_2(\beta), && (9)\\ Q_\beta(\beta) &= -m_2(\beta)' W m_2(\beta), && (10) \end{aligned}$$

where $E[T^{-1}\sum_{t=1}^{T}\phi_t(\theta)] = m_1(\theta)/\sqrt{T}+m_2(\beta)+o(1)$, $m_1(\theta)$ 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. Under the alternative hypothesis, parameters can be all unidentified, i.e., $\alpha = \theta$, $\beta = \emptyset$, $k_1 = k$ and $k_2 = 0$.

4. 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.

5. 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.

6. 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).

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

(a) Under the null hypothesis,

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

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

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

$$\hat{d}_T \;=\; T^{\frac{1}{2}}(\tilde{\theta}_{2T} - \tilde{\theta}_{1T} - \lambda_T\hat{B}_{2T} + \lambda_T\hat{B}_{1T}),$$

$$\hat{D}_T' \;=\; \begin{pmatrix} -[\nabla_{\theta\theta}Q_{1T}(\tilde{\theta}_{1T}) - \lambda_T I_k]^{-1} & [\nabla_{\theta\theta}Q_{2T}(\tilde{\theta}_{2T}) - \lambda_T I_k]^{-1} \end{pmatrix}, \tag{20}$$

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

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 $\hat{\theta}_1$ and $\hat{\theta}_2$, 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

$$D' \;=\; \begin{pmatrix} -[\nabla_{\theta\theta}Q_1(\theta_0)]^{-1} & [\nabla_{\theta\theta}Q_2(\theta_0)]^{-1} \end{pmatrix}, \tag{22}$$

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

(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 noncentral χ^2 random function of $\alpha \in \Theta_A$:

$$N\left(E(z_i z_i')C(\alpha-\alpha_0), E(\varepsilon_i-(\alpha-\alpha_0)\eta_i)^2 E(z_i z_i')\right)' \times N\left(E(z_i z_i')C(\alpha-\alpha_0), E(\varepsilon_i-(\alpha-\alpha_0)\eta_i)^2 E(z_i z_i')\right), \quad (25)$$

   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 noncentral χ^2 random function of $\alpha \in \Theta_A$:

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

   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. 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 heteroskedasticity 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 a 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.[3] In empirical macroeconomics we generally have a plenty of candidates for Z_2 such as lagged values of Z_1.

4. In linear IV models, Hahn, Ham and Moon (2008) show that the conventional Hausmann test is invalid when instruments are weak. If different sets of instruments are used to construct $\hat{\theta}_{1T}$ and $\hat{\theta}_{2T}$, e.g. X_1 and $X = [X_1, X_2]$, our test will reject the null hypothesis with probability approaching one asymptotically when parameters are strongly identified and X_2 is endogenous. If the purpose is to test for weak identification, one should not construct Q_{1T} and Q_{2T} using different instruments or moment conditions.

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

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

$$\hat{\theta}_{1,T} = (Y'X\hat{W}_T X'Y)^{-1}Y'X\hat{W}_T X'y, \tag{27}$$

$$\hat{\theta}_{2,T} = (Y'XX'Y)^{-1}Y'XX'y, \tag{28}$$

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

$$\begin{bmatrix} Y'X\hat{W}_T X_i(y_i - \hat{\theta}_{1,T}'X_i) \\ Y'XX_i(y_i - \hat{\theta}_{2,T}'X_i) \end{bmatrix}. \tag{29}$$

6. 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

$$\begin{bmatrix} \sum_{s=1}^{T} D_\theta\phi_s(\hat{\theta}_{1,T})\hat{W}_T\phi_t(\hat{\theta}_{1,T}) \\ \sum_{s=1}^{T} D_\theta\phi_s(\hat{\theta}_{1,T})'\hat{W}_T\phi_t(\hat{\theta}_{2,T}) \end{bmatrix}, \tag{30}$$

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)–(h) hold.

(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,

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

where

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

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 the test rejects the null hypothesis with probability approaching one whether parameters are not identified at all or only weakly identified.

3. 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 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. In what follows, we provide a step-by-step description of

[4]See equation (45). When multiplied by $T^{1/2}$ there is error of order $O_p(\lambda_T)$.

how to implement our test by choosing the shrinkage parameter via a cross-validation procedure.[5] We investigate the small sample properties of the procedure in the next section.

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 $Q_{1T}(\theta) = \bar{m}_T(\theta)'W_{1T}\bar{m}_T(\theta)$, $Q_{2T}(\theta) = \bar{m}_T(\theta)'W_{1T}\bar{m}_T(\theta)$, $\bar{m}_T(\theta) = (1/T)\sum_{t=1}^{T} m(z_t,\theta)$ 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 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 $Q_{1T,t}(\theta) = \bar{m}_{T,t}(\theta)'W_{1T}\bar{m}_{T,t}(\theta)$, $Q_{2T,t}(\theta) = \bar{m}_{T,t}(\theta)'W_{1T}\bar{m}_{T,t}(\theta)$ 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 Mean Squared error estimate of these parameter estimates:

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

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

[5] See Carrasco (2008, Section 4) for cross-validation methods.

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

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

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

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.

4 Monte Carlo Experiments

We analyze the finite sample performance of the test that we propose 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.

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

$$\begin{bmatrix} \log\left(\frac{C_t}{C_{t-1}}\right) \\ \log\left(\frac{D_t}{D_{t-1}}\right) \end{bmatrix} = \mu + \Phi \begin{bmatrix} \log\left(\frac{C_{t-1}}{C_{t-2}}\right) \\ \log\left(\frac{D_{t-1}}{D_{t-2}}\right) \end{bmatrix} + \begin{bmatrix} u_{ct} \\ u_{dt} \end{bmatrix}, \tag{34}$$

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

$$E_t \left[\delta R_{t+1} \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} - 1 \right] = 0, \tag{35}$$

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 where $\theta = [\delta, \gamma]'$, and $\Theta = [0.7, 1.3] \times [0, 30]$. In our notation the objective function can be written as

$$Q_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \left[\delta R_{t+1} \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} - 1 \right] Z_t' W_T \frac{1}{T} \sum_{t=1}^{T} Z_t \left[\delta R_{t+1} \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} - 1 \right], \tag{36}$$

$Z_t = [1, R_t, C_t/C_{t-1}]'$, 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 $K = \{1, 5, 10\}$ in this Monte Carlo experiment. Unlike simple parametric hypothesis, 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 larger 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. 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. The size of Wright's (2003) test is distorted when the parameters are weakly identified.

4.2 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:[6]

$$E_t \left[\left\{ r_t - \left[rr^* - (\beta - 1) \pi^* + \beta \pi_{t+1} + \gamma y_{t+1} \right] \right\} X_t \right] = 0 \tag{37}$$

where r_t is the Fed Fund Rate, π_{t+1} is the inflation rate, and y_{t+1} is the average output gap between time t and $t+1$. Let $Z_t = \{\pi_{t+1}, y_{t+1}, 1\}'$, where $Z_t \sim N_{3\times 1}(0, \Omega_Z)$ and Ω_Z is set to be a positive definite matrix chosen randomly and equal to [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]. 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 iid(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 $rr^* - (\beta - 1) \pi^* = 1$ (which means that on average the Central Bank aims at zero inflation).

The set of instruments consists of a constant as well as:

$$X_t = B'_{xz} Z_t + u_{X,t}$$

where $B_{xz} \equiv \vartheta \left[I_{2\times 2} \;\; 0_{2\times 2} \right]$ and $u_{X,t} \sim N_{2\times 1}(0_{2\times 1}, I_{4\times 4})$. 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.[7] 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.

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) has a

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

[7] 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.

tendency to reject the null hypothesis of no identification when the parameters are weakly identified. Our test, implemented with the 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.

5 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 nonlinear 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 the prior rather than by 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 (2006) and Ruge-Murcia (2005) 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). Clarida, Gali and Gertler (2000, hereafter CGG) estimate the monetary policy reaction function by GMM based on the following moment conditions:

$$E_t\left[\left\{r_t-(1-\rho_1-\rho_2)\left[rr^*-(\beta-1)\pi^*+\beta\pi_{t+1}+\gamma y_{t+1}\right]-\rho_1 r_{t-1}-\rho_2 r_{t-2}\right\}X_t\right]=0 \tag{38}$$

where r_t is the Fed Fund Rate (FYFF from CITIBASE), π_{t+1} is the inflation rate (the rate of change of prices between time t and $t+1$ expressed in annual rates), and y_{t+1} is the average output gap at time $t+1$ (where the output gap is defined as the deviation between the actual GDP and its potential, either from CBO or detrended GDP or unemployment). 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 (38); it is instead estimated as: $\left[(\widehat{rr^*-(\beta-1)\pi^*})-\widehat{rr^*}\right]/(1-\beta)$, where 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.

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

$$E_t\left[\left\{r_t-\alpha_1-\alpha_2\pi_{t+1}-\alpha_3 r_{t-1}-\alpha_4 r_{t-2}-\alpha_5 y_{t+1}\right\}X_t\right]=0 \tag{39}$$

that is, $E\left[g_t(\alpha)\right]=0$, where $g_t(\alpha)\equiv(r_t-\alpha' Z_t)X_t$ 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 $Q_T(\alpha)=-\frac{1}{2}\overline{g}_T(\alpha)' W\overline{g}_T(\alpha)$, where $\overline{g}_T(\alpha)=T^{-1}\sum_{t=1}^T g_t(\alpha)=T^{-1}\sum_{t=1}^T X_t r_t - T^{-1}\sum_{t=1}^T X_t Z_t'\alpha$, $G=T^{-1}\sum_{t=1}^T \partial g_t(\alpha)/\partial\alpha' = -T^{-1}\sum_{t=1}^T X_t Z_t'$, and $\nabla_{\theta\theta}Q_T(\theta,W)=-G'WG$.

The shrinkage GMM estimator satisfies:

$$\begin{aligned}\widetilde{\alpha}(W) &= \arg\max_{\alpha}\left(Q_T(\alpha)-0.5\lambda||\alpha||^2\right) \qquad (40)\\ &= \arg\max_{\alpha}\left(-\frac{1}{2}\overline{g}_T(\alpha)' W\overline{g}_T(\alpha)-0.5\lambda_T\sum_{s=1}^5 {\alpha_s}^2\right)\end{aligned}$$

From (40), the first order conditions give:

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

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 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 (2007) estimates the joint confidence sets for the inflation-aversion and output gap reaction parameters by using Stock and Wright's (2000) identification-robust test.[8] 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.

6 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).

[8] 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.

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 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.

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

$$\nabla_\theta Q_T(\tilde{\theta}_T) - \lambda_T(\tilde{\theta}_T - \bar{\theta}) \;=\; 0_{k\times 1}, \tag{41}$$

By applying the mean value theorem to (41) we obtain

$$\tilde{\theta}_T - \theta_0 - \lambda_T[\nabla_{\theta\theta} Q_T(\bar{\theta}_T) - \lambda_T I_k]^{-1}\left(\theta_0 - \bar{\theta}\right) \;=\; -[\nabla_{\theta\theta} Q_T(\bar{\theta}_T) - \lambda_T I_k]^{-1}\nabla_\theta Q_T(\theta_0), \tag{42}$$

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 nonsingular with probability approaching one. Thus,

$$\begin{aligned} & \frac{\partial}{\partial\theta}\text{vec}\left\{[\nabla_{\theta\theta} Q_T(\theta) - \lambda_T I_k]^{-1}\right\} \\ = \quad & -\left\{[\nabla_{\theta\theta} Q_T(\theta) - \lambda_T I_k]^{-1} \otimes [\nabla_{\theta\theta} Q_T(\theta) - \lambda_T I_k]^{-1}\right\}\frac{\partial}{\partial\theta}\text{vec}[\nabla_{\theta\theta} Q_T(\theta)] \end{aligned} \tag{43}$$

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

$$[\nabla_{\theta\theta} Q_T(\bar{\theta}_T) - \lambda_T I_k]^{-1} \;=\; [\nabla_{\theta\theta} Q_T(\theta_0) - \lambda_T I_k]^{-1} + O_p(\|\tilde{\theta}_T - \theta_0\|). \tag{44}$$

It follows from (42) and (44) that

$$\begin{aligned} \tilde{\theta}_T - \theta_0 - \lambda_T B_T(\theta_0) \;&=\; -[\nabla_{\theta\theta} Q_T(\theta_0) - \lambda_T I_k]^{-1}\nabla_\theta Q_T(\theta_0) \\ &\quad + O_p(\lambda_T\|\tilde{\theta}_T - \theta_0\|) + O_p(\lambda_T\|\tilde{\theta}_T - \theta_0\|). \end{aligned} \tag{45}$$

Therefore equation (12) follows from (45) 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.iv) that

$$Q_T(\theta) \stackrel{p}{\to} Q_\beta(\beta) \tag{46}$$

uniformly in θ. Because $Q_\beta(\beta)$ is uniquely maximized at β_0 by Assumption (d.iv), we can show that $\hat{\beta}_T \stackrel{p}{\to} \beta_0$ 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

$$\nabla_\beta Q_T(\hat{\theta}_T) \;=\; 0_{k_2 \times 1}. \tag{47}$$

By applying the mean value theorem to (47) we obtain

$$\nabla_\beta Q_T(\theta_0) + \nabla_{\beta\alpha} Q_T(\bar{\theta}_T)(\hat{\alpha}_T - \alpha_0) + \nabla_{\beta\beta} Q_T(\bar{\theta}_T)(\hat{\beta}_T - \beta_0) \;=\; 0_{k_2 \times 1}, \tag{48}$$

where $\bar{\theta}_T = [\bar{\alpha}_T', \bar{\beta}_T']'$ is a point between θ_0 and $\hat{\theta}_T$. Because $\hat{\beta}_T \stackrel{p}{\to} \beta_0$, Θ is compact by Assumption (a), $\nabla_{\beta\alpha} Q_T(\theta) = O_p(T^{-1/2})$ 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 nonsingular by Assumptions (a), (b) and (c.iii) we have

$$\hat{\beta}_T - \beta_0 \;=\; O_p(T^{-1/2}). \tag{49}$$

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

$$\begin{aligned} & TQ_T(a, \beta_0 + bT^{-1/2}) \\ = \;& TQ_T(a, \beta_0) + T^{\frac{1}{2}} \nabla_\beta Q_T(a, \beta_0) b + \frac{1}{2T} b' \nabla_{\beta\beta} Q_T(a, \beta_0) b + o_p(T^{-1}) \\ \Rightarrow \;& Q_\alpha(a, \beta_0) + Z_\beta(\alpha, \beta_0)' b + \frac{1}{2} b' \nabla_{\beta\beta} Q_\beta(\beta_0)(a, \beta_0) b. \end{aligned} \tag{50}$$

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 $\tilde{\alpha}_T$,

$$\nabla_\alpha Q_T(\tilde{\alpha}_T, \tilde{\beta}_T) - \lambda_T(\tilde{\alpha}_T - \bar{\alpha}) \;=\; 0_{k_1 \times 1}, \tag{51}$$

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 $\tilde{\beta}_T$,

$$\nabla_\beta Q_T(\tilde{\alpha}_T, \tilde{\beta}_T) - \lambda_T(\tilde{\beta}_T - \bar{\beta}) \;=\; 0_{k_1 \times 1}, \tag{52}$$

around β_0 yields

$$\begin{aligned} & \tilde{\beta}_T - \beta_0 - \lambda_T[\nabla_{\beta\beta} Q_T(\tilde{\alpha}_T, \bar{\beta}_T) - \lambda_T I_{k_2}]^{-1}(\beta_0 - \bar{\beta}) \\ = \;& -[\nabla_{\beta\beta} Q_T(\tilde{\alpha}_T, \bar{\beta}_T) - \lambda_T I_{k_2}]^{-1} \nabla_\beta Q_T(\tilde{\alpha}_T, \beta_0), \end{aligned} \tag{53}$$

where $\bar{\beta}_T$ is a point between $\tilde{\beta}_T$ and β_0. By Assumptions (d.i), (d.iv) and (e), (53) can be written as

$$\begin{aligned} & \tilde{\beta}_T - \beta_0 - \lambda_T[\nabla_{\beta\beta} Q_\beta(\beta_0)]^{-1}(\beta_0 - \bar{\beta}) \\ = \;& -[\nabla_{\beta\beta} Q_\beta(\beta_0)]^{-1} \nabla_\beta Q_T(\bar{\alpha}, \beta_0) + O_p(T^{-\frac{1}{2}} \|\tilde{\alpha}_T - \overline{\alpha}\|) + O_p(T^{-\frac{1}{2}}). \end{aligned} \tag{54}$$

It follows from (54) and Assumptions (a), (d.i), (d.iv) and (e) that

$$\tilde{\beta}_T - \beta_0 = O_p(T^{-1/2}). \tag{55}$$

It follows from (51), (55) and Assumptions (d.ii) and (d.iii) that

$$\begin{aligned} \tilde{\alpha}_T - \bar{\alpha} \;&=\; \frac{1}{\lambda_T} \nabla_\alpha Q_T(\tilde{\alpha}_T, \beta_0) + \frac{1}{\lambda_T} \nabla_{\alpha\beta} Q_T(\tilde{\alpha}_T, \bar{\beta}_T)(\tilde{\beta}_T - \beta_0) \\ &=\; O_p\left(\frac{1}{\lambda_T T}\right), \end{aligned} \tag{56}$$

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

Second we will consider a limiting representation for

$$\begin{aligned} & Q_T\left(\bar{\alpha} + \frac{a}{\lambda_T T}, \beta_0 + \lambda_T[\nabla_{\beta\beta} Q_\beta(\beta_0)]^{-1}(\beta_0 - \bar{\beta}) + T^{-\frac{1}{2}} b\right) \\ & + \frac{\lambda_T}{2}\left(\left\|\frac{a}{\lambda_T T}\right\|^2 + \|\beta_0 + \lambda_T[\nabla_{\beta\beta} Q_\beta(\beta_0)]^{-1}(\beta_0 - \bar{\beta}) + T^{-\frac{1}{2}} b - \bar{\beta}\|^2\right) \end{aligned} \tag{57}$$

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 (57) can be written as

$$\begin{aligned}
& Q_T(\bar{\alpha}, \beta_0) + \frac{\lambda_T}{2}\|\beta_0 - \bar{\beta}\|^2 \\
& + \left[\nabla_\theta Q_T(\bar{\alpha}, \beta_0) - \lambda_T \begin{pmatrix} 0_{k_1 \times 1} \\ \beta_0 - \bar{\beta} \end{pmatrix}\right]' \begin{bmatrix} \frac{a}{\lambda_T T} \\ \lambda_T[\nabla_{\beta\beta} Q_\beta(\beta_0)]^{-1}(\beta_0 - \bar{\beta}) + T^{-\frac{1}{2}} b \end{bmatrix} \\
& + \frac{1}{2} \begin{bmatrix} \frac{a}{\lambda_T T} \\ \lambda_T[\nabla_{\beta\beta} Q_\beta(\beta_0)]^{-1}(\beta_0 - \bar{\beta}) + T^{-\frac{1}{2}} b \end{bmatrix}' [\nabla_{\theta\theta} Q_T(\bar{\alpha}_T, \bar{\beta}_T) - \lambda_T I_k] \\
& \times \begin{bmatrix} \frac{a}{\lambda_T T} \\ \lambda_T[\nabla_{\beta\beta} Q_\beta(\beta_0)]^{-1}(\beta_0 - \bar{\beta}) + T^{-\frac{1}{2}} b \end{bmatrix}, \qquad (58)
\end{aligned}$$

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 (58) and Assumptions (d) and (e) that

$$\begin{aligned}
& T\left[Q_T\left(\bar{\alpha} + \frac{a}{\lambda_T T}, \beta_0 + \lambda_T[\nabla_{\beta\beta} Q_\beta(\beta_0)]^{-1}(\beta_0 - \bar{\beta}) + T^{-\frac{1}{2}} b\right)\right. \\
& + \frac{\lambda_T}{2}\left(\left\|\frac{a}{\lambda_T T}\right\|^2 + \|\beta_0 + \lambda_T[\nabla_{\beta\beta} Q_\beta(\beta_0)]^{-1}(\beta_0 - \bar{\beta}) + T^{-\frac{1}{2}} b - \bar{\beta}\|^2\right) \\
& \left. - Q_T(\bar{\alpha}, \beta_0) + \frac{\lambda_T}{2}\|\beta_0 - \bar{\beta}\|^2\right] \\
\Rightarrow \; & [Z_\beta(\bar{\alpha}, \beta_0) - \kappa(\beta_0 - \bar{\beta})]'(\kappa[\nabla_{\beta\beta} Q_\beta(\beta_0)]^{-1}(\beta_0 - \bar{\beta}) + b) \\
& + \frac{1}{2}(\kappa[\nabla_{\beta\beta} Q_\beta(\beta_0)]^{-1}(\beta_0 - \bar{\beta}) + b)' \nabla_{\beta\beta} Q_\beta(\beta_0)(\kappa[\nabla_{\beta\beta} Q_\beta(\beta_0)]^{-1}(\beta_0 - \bar{\beta}) + b). \\
& \qquad (59)
\end{aligned}$$

Therefore Theorem 1(b) follows from Lemma 3.2.1 of van der Vaart and Wellner (1996, p.286), (56), (59) 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 (45) that

$$\begin{aligned}
& T^{\frac{1}{2}}[\tilde{\theta}_{2T} - \tilde{\theta}_{1T} - \lambda_T \hat{B}_{2T} + \lambda_T \hat{B}_{1T}] \\
= & -T^{\frac{1}{2}}[\nabla_{\theta\theta} Q_{2T}(\theta_0) - \lambda_T I_k]^{-1} \nabla_\theta Q_{2T}(\theta_0) + T^{\frac{1}{2}}[\nabla_{\theta\theta} Q_{1T}(\theta_0) - \lambda_T I_k]^{-1} \nabla_\theta Q_{1T}(\theta_0) \\
& -\lambda_T T^{\frac{1}{2}}(\hat{B}_{2T} - B_{2T}) + \lambda_T T^{\frac{1}{2}}(\hat{B}_{1T} - B_{1T}) + O_p(T^{-\frac{1}{2}}) \\
\xrightarrow{d} & N(0_{k\times 1}, D'\Sigma D). \qquad (60)
\end{aligned}$$

Since $\hat{D}_T \xrightarrow{p} D$ and $\hat{\Sigma}_T \xrightarrow{p} \Sigma$ by Assumptions (c.iii), (e) and (g) and Theorem 1(a), the desired result follows from (60).

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

$$\begin{aligned}
& [\nabla_{\theta\theta} Q_{jT}(\theta) - \lambda_T I_k]^{-1} \\
= & \begin{bmatrix} T^{-1} H_{j,\alpha\alpha}(\theta) - \lambda_T I_{k_1} + o_p(T^{-1}) & T^{-1/2} H_{j,\alpha\beta}(\theta) + o_p(T^{-1/2}) \\ T^{-1/2} H_{j,\beta\alpha}(\theta) + o_p(T^{-1/2}) & \nabla_{\beta\beta} Q_{j,\beta}(\beta) - \lambda_T I_{k_2} + O_p(T^{-1/2}) \end{bmatrix}^{-1} \\
= & \begin{bmatrix} -\frac{1}{\lambda_T} I_{k_1} + O_p\left(\frac{1}{\lambda_T^2 T}\right) & O_p\left(\frac{1}{\lambda_T T^{1/2}}\right) \\ O_p\left(\frac{1}{\lambda_T T^{1/2}}\right) & [\nabla_{\beta\beta} Q_{j,\beta}(\beta) - \lambda_T I_{k_2}]^{-1} + O_p(T^{-\frac{1}{2}}) \end{bmatrix}, \qquad (61)
\end{aligned}$$

and

$$\lambda_T \hat{B}_{jT} = \begin{bmatrix} -\alpha_j^* + \bar{\alpha} + O_p(\frac{1}{\lambda_T T}) \\ \lambda_T [\nabla_{\beta\beta} Q_{j,\beta}(\beta_0) - \lambda_T I_{k_2}]^{-1} (\beta_0 - \bar{\beta}) + O_p(\frac{\lambda_T}{T^{1/2}}) \end{bmatrix}. \qquad (62)$$

It follows from Theorem 1(b) and equations (61) and (62) that

$$\hat{d}_T = \tilde{\theta}_{2,T} - \tilde{\theta}_{1,T} - \lambda_T \hat{B}_{2,T} + \lambda_T \hat{B}_{1,T} \Rightarrow \begin{bmatrix} \alpha_2^* - \alpha_1^* \\ 0_{k_2\times 1} \end{bmatrix} \qquad (63)$$

and that

$$\begin{aligned}
& \left[I_2 \otimes \begin{pmatrix} T^{-\frac{1}{2}} I_{k_1} & 0_{k_1\times k_2} \\ 0_{k_2\times k_1} & I_{k_2} \end{pmatrix} \right] \hat{D}_T \\
\xrightarrow{d} & \begin{pmatrix} -\begin{pmatrix} -\frac{1}{\kappa} I_{k_1} & 0_{k_1\times k_2} \\ \frac{1}{\kappa}(\nabla_{\beta\beta} Q_{1,\beta}(\beta_0))^{-1} H_{1,\beta\alpha}(\bar{\alpha}, \beta_0) & (\nabla_{\beta\beta} Q_{1,\beta}(\beta_0))^{-1} \end{pmatrix} \\ \begin{pmatrix} -\frac{1}{\kappa} I_{k_1} & 0_{k_1\times k_2} \\ \frac{1}{\kappa}(\nabla_{\beta\beta} Q_{2,\beta}(\beta_0))^{-1} H_{2,\beta\alpha}(\bar{\alpha}, \beta_0) & (\nabla_{\beta\beta} Q_{2,\beta}(\beta_0))^{-1} \end{pmatrix} \end{pmatrix} \\
\equiv & \left[I_2 \otimes \begin{pmatrix} \frac{1}{\kappa} I_{k_1} & 0_{k_1\times k_2} \\ 0_{k_2\times k_1} & I_{k_2} \end{pmatrix} \right] \begin{bmatrix} -M_1 \\ M_2 \end{bmatrix} \qquad (64)
\end{aligned}$$

By Assumption (h.ii),

$$\begin{aligned}
&\left[I_2 \otimes \begin{pmatrix} T^{\frac{1}{2}} I_{k_1} & 0_{k_1 \times k_2} \\ 0_{k_2 \times k_1} & I_{k_2} \end{pmatrix}\right]' \hat{\Sigma}_T \left[I_2 \otimes \begin{pmatrix} T^{\frac{1}{2}} I_{k_1} & 0_{k_1 \times k_2} \\ 0_{k_2 \times k_1} & I_{k_2} \end{pmatrix}\right] \\
\Rightarrow &\begin{bmatrix} \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{21}^* & \Sigma_{22}^* \end{bmatrix} \qquad (65)
\end{aligned}$$

The result (32) follows from equations (63), (64) and (65) and equa2tion (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

1. An, Sungbae, and Frank Schorfheide (2007), "Bayesian Analysis of DSGE Models", *Econometric Reviews* 26(2-4), 113-172.

2. Canova, Fabio, and Luca Sala (2006), “Back to Square One: Identification Issues in DSGE Models,” unpublished manuscript, *University of Pompeu Fabra and Universitá Bocconi.*

3. Carrasco, Marine (2008), “A Regularization Approach to the Many Instruments Problem,” unpublished manuscript, *Université de Montréal.*

4. Choi, In, and Peter C.B. Phillips (1992), “Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations,” *Journal of Econometrics,* 51, 113–150.

5. Clarida, Richard, Jordi Galí, Mark Gertler (2000), “Monetary policy rules and macroeconomic stability: evidence and some theory”, *Quarterly Journal of Economics* 115, 147–180.

6. Cragg, John G., and Stephen 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.

7. Cragg, John G., and Stephen G. Donald (1997), “Inferring the Rank of a Matrix,” *Journal of Econometrics* 76, 223–250.

8. Dufour, Jean-Marie (1997), “Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models,” *Econometrica,* 65, 1365–1387.

9. Dufour, Jean-Marie, Lynda Khalaf, and Maraf Kichian (2006), “Inflation Dynamics and the New Keynesian Phillips Curve: An Identification Robust Econometric Analysis,” *Journal of Economic Dynamics and Control*, 30, 1707–1727.

10. Gallant, Ronald A., and Halbert White (1988), *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models,* Basil Blackwell: Oxford, UK.

11. Gill, Len, and Arthur 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.

12. Guggenberger, Patrick, and Richard J. Smith (2005), "Generalized Empirical Likelihood Estimators and Tests under Partial, Weak and Strong Identification," *Econometric Theory,* 21, 667–709.

13. Hahn, Jinyong, and Jerry Hausman (2002), "A New Specification Test for the Validity of Instrumental Variables," *Econometrica,* 70, 163–189.

14. Hahn, Jinyong, John Ham and Hyungsik Roger Moon (2008), "The Hausman Test and Weak Instruments," unpublished manuscript, *UCLA, USC and USC.*

15. Hausmann, Jerry, James H. Stock and Motohiro Yogo (2005), "Asymptotic Properties of the Hahn-Hausman Test for Weak Instruments," *Economics Letters*, 89, 333–342.

16. Hoerl, Arthur E., and Robert W. Kennard (1970a), "Ridge Regression: Biased Estimation for Non-orthogonal Problems," *Technometrics,* 12, 55–67.

17. Hoerl, Arthur E., and Robert W. Kennard (1970b), "Ridge Regression: Applications to Non-orthogonal Problems," *Technometrics,* 12, 69–82.

18. Horn, Roger A., and Charles R. Johnson (1985), *Matrix Analysis,* Cambridge University Press: Cambridge, UK.

19. Iskrev, Nikolay (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.*

20. Kleibergen, Frank, and Richard Paap (2006), "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics* 133, 97–126.

21. Kocherlakota, Narayana (1990), “On Tests of Representative Consumer Asset Pricing Models,” *Journal of Monetary Economics,* 26, 285–304.

22. Liu, Ta-Chung (1960), “Under-identification, Structural Estimation, and Forecasting,” *Econometrica,* 28, 855–865.

23. Magnus, Jan R., and Heinz Neudecker (1999), *Matrix Differential Calculus with Applications in Statistics and Econometrics,* Revised Edition, John Wiley & Sons: Chichester, UK.

24. Mavroeidis, Sophocles (2006), “Monetary Policy Rules and Macroeconomic Stability: Some New Evidence,” unpublished manuscript, *Brown University.*

25. Nason, James M., and Gregor W. Smith (2005), “Identifying the New Keynesian Phillips Curve,” *Federal Reserve Board of Atlanta*, Working Paper 2005-1.

26. Newey, Whitney K., and D.M. McFadden (1994), “Large Sample Estimation and Hypothesis Testing,” in Robert F. Engle and Daniel L. McFadden eds. *the Handbook of Econometrics,* Volume IV, Elsevier: Amsterdam, the Netherlands.

27. Okui, Ryo (2007), “Instrumental Variables Estimation in the Presence of Many Moment Conditions,” unpublished manuscript, *Hong Kong University of Science and Technology.*

28. Ruge-Murcia, Francisco (2007), ”Methods to Estimate Dynamic Stochastic General Equilibrium Models”, *Journal of Economic Dynamics and Control* 31, 2599-2636.

29. Smets, Frank, and Rafael Wouters (2005), ”Shocks and Frictions in U.S. Business Cycles: A Bayesian DSGE Approach”, *The American Economic Review* 97(3), 586-606.

30. Smith, Richard J., and Jean-Marc Robins (2000), “Tests of Rank,” *Econometric Theory,* 16, 151–175.

31. Staiger, Douglas, and James H. Stock (1997), “Instrumental Variables Regressions with Weak Instruments,” *Econometrica,* 65, 557–586.

32. Stock, James H. and Motohiro 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, 80–108.

33. Stock, James H. and Jonathan H. Wright (2000), "GMM With Weak Identification," *Econometrica,* 68, 1055–1096.

34. Stock, James H., Jonathan H. Wright, and Motohiro Yogo (2005), "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics,* 20, 518–529.

35. Tauchen, George 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.

36. Tauchen, George E., and Robert Hussey (1991), "Quadrature Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models," *Econometrica,* 59, 371–396.

37. Taylor, John B. (1993), "Discretion versus Policy Rules in practice", *Carnegie-Rochester Conference Series on Public Policy* 39, 195-214.

38. Yogo, Motohiro, (2004), "Estimating the Elasticity of Intertemporal Substitution When Instruments are Weak," *Review of Economics and Statistics,* 86, 797–810.

39. Van der Vaart, Aad W., and Jon A. Wellner (1996), *Weak Convergence and Empirical Processes With Applications to Statistics,* Springer-Verlag: New York, NY.

40. Wright, Jonathan H. (2002), "Testing the Null Identification in GMM," *Board of Governors of the Federal Reserve System,* International Finance Discussion Paper Number 732.

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

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	Ours
Model	T	bias	coverage	bias	coverage	(2003)	
SI	50	0.007	0.927	0.138	0.892	0.744	0.070
	100	0.005	0.936	0.099	0.917	0.906	0.046
	200	0.003	0.944	0.070	0.937	0.975	0.035
PI1	50	0.034	0.986	1.887	0.996	0.136	0.900
	100	0.034	0.992	1.903	0.999	0.115	0.945
	200	0.032	1.000	1.855	0.999	0.097	0.952
PI2	50	0.138	0.417	12.188	0.244	0.129	0.941
	100	0.138	0.415	12.156	0.248	0.112	0.988
	200	0.137	0.411	12.197	0.250	0.100	0.999
WI1	50	0.012	0.992	0.612	0.682	0.319	0.558
	100	0.008	0.994	0.599	0.676	0.308	0.664
	200	0.005	0.984	0.532	0.687	0.358	0.770
WI2	50	1.032	0.124	14.488	0.768	0.459	0.962
	100	1.139	0.164	13.203	0.752	0.349	0.925
	200	0.719	0.912	3.144	0.992	0.266	0.995

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 two columns).

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

T	Model	Wright (2003)	Our Test
50	SI	1	0.08
	WI	0.19	0.52
	NI	0.04	0.69
100	SI	1	0.06
	WI	0.26	0.58
	NI	0.04	0.76
200	SI	1	0.05
	WI	0.41	0.62
	NI	0.07	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").

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.