Identifying Euler equation models via stability restrictions

Leandro M. Magnusson
Tulane University

Sophocles Mavroeidis
Brown University

April 1, 2009

Preliminary and incomplete. Please do not cite.

Abstract

Structural models that are immune to the Lucas (1976) critique have stable parameters in the face of breaks induced by policy regime shifts. We show that such breaks represent exogenous variation that can be usefully exploited for structural inference and that this information is ignored by standard methods that rely only on the average variation in the data over the sample. We propose efficient methods for exploiting this information that do not require any assumptions about identification, specification of the reduced-form dynamics, existence of breaks or knowledge of the break dates. Application of these methods to the two core equations of the new Keynesian policy model provides substantial improvement to the identification of the models' parameters.

Keywords: GMM, identification, structural stability, Lucas critique, new Keynesian models.

JEL: C22, E31

1 Introduction

An important objective of structural macroeconomic models is to characterize deep structural relationships that are invariant to changes in the distribution of the data. The parameters of such models admit a structural interpretation and are assumed to be stable over time. Models that do not possess that property are subject to the well-known Lucas (1976) critique. Lucas (1976) pointed out that econometric models will break down when the underlying economic environment changes, for example because of policy shifts, unless these models adequately account for agents' reaction to these changes. In econometric terms, the parameters of models that are immune to the Lucas critique should remain invariant to exogenous changes in the data generating process.

The contribution of this paper is twofold. First, we show that stability restrictions (i.e., immunity to the Lucas critique) constitute an important and powerful source of identification in Euler equation models. The key insight is that changes in the distribution of the data induced by, for example, policy regime shifts, provide additional exogenous variation that can be usefully exploited for inference. The usual estimation approach relies only on the identifying assumption that certain moment restrictions hold *on average* over the full sample, and this ignores subsample variation in the data. We show that this approach can only be justified when there are no changes in the data generating process. We think this assumption is too strong in many context, especially in macroeconomics, where there is considerable evidence of parameter instability, see, for example, Stock and Watson (1996), Clarida, Galí, and Gertler (2000) and Sims and Zha (2006). So, we expect *a priori* that the information contained in stability restrictions will be nontrivial, and our applications confirm this empirically.

It is common practice to use stability restrictions exclusively for post-estimation misspecification testing by means of the various parameter stability or structural change tests, that have been proposed in the literature, see Andrews (1993), Andrews and Ploberger (1994), Elliott and Mueller (2006), or Perron (2005) and the references therein. When the null hypothesis of parameter stability is rejected, the Lucas critique applies and the validity of the model is put into question. Thus, to the best of our knowledge, stability restrictions are being used exclusively to detect weaknesses of a model. We show that this approach does not make efficient use of the information in the data, and we propose instead to use stability restrictions constructively in order to sharpen inference on the parameters of the model.

The second contribution of this paper is to propose econometric tools that can be used to extract the information in the stability restrictions for structural inference. The methods we propose require only mild assumptions about the nature of instability in the distribution of the data. It should be stressed that these assumptions are weaker than those needed to justify the standard parameter stability tests mentioned above, and therefore the scope of the proposed methods is very wide. Specifically, they do not require any assumptions about identification of the structural parameters, nor any prior knowledge about the incidence, number and timing of breaks in reduced-form parameters. Our analysis is carried out in the framework of the Generalized Method of Moments (GMM), which is suitable for estimating Euler equation models. Our proposed methods can be used to obtain confidence sets for the structural parameters that have the interpretation that they contain all the values of the parameters (i.e., all possible 'structures') for which the model's identifying restrictions hold in every subsample and the model is immune to the Lucas critique.

We examine the empirical relevance of the proposed methods by applying them to two equations that form the core of the new Keynesian macroeconomic policy model: the new Keynesian Phillips curve, and the Euler equation for output, which is sometimes also referred to as the new Keynesian IS curve. These models are well-known to suffer from problems of weak instruments, see Kleibergen and Mavroeidis (2009) and Fuhrer and Rudebusch (2004) for the inflation and output models, respectively. The parameter that is typically most difficult to estimate accurately is the coefficient that governs the degree of forward-looking behavior, or rational versus adaptive expectations. Identification-robust 90% confidence intervals for this parameter that use only full-sample information contain the entire parameter space. However, using methods that exploit the stability restrictions, we find that the confidence sets on the parameters become smaller, and these results are robust to alternative datasets and sample periods. Moreover, the results indicate that the most commonly used versions of the models are immune to the Lucas critique.

More specifically, the improvement in the identification of the parameters is most evident when we use the Federal Funds rate as an instrument, and it is stronger for the output gap equation than for the Phillips curve. This finding is consistent with the view that monetary policy has not been stable over the postwar period, and could be viewed as indirect evidence on this issue. The most striking results are obtained for the Euler equation for output: the fraction of forward-looking agents is estimated fairly precisely around 0.5, and the elasticity of substitution, which is a key parameter in the monetary transmission mechanism, is tightly estimated around zero. We note that these results are comparable to estimates reported by Fuhrer and Rudebusch (2004) using full-information maximum likelihood methods, even though our methods do not require the specification of the dynamics of the real interest rate and inflation, nor any assumptions about identification. Hence, these findings show that we can make substantial progress in identifying the structural parameters in Euler equation models without imposing additional restrictions that typically reduce the scope of the results.

The outline of the paper is as follows. Section 2 discusses the main idea of the paper in the context of a simple stylized example. Section 3 proposes econometric methods that can be used to exploit that idea, and section 4 reports simulations on their size and power properties. The empirical applications are given in Section 5, and 6 concludes.

2 Identification by stability restrictions

To fix ideas, we start with a very simple stylized example, which is sufficient to present the main intuition of the results. This example can be thought of as a stripped down version of some more realistic dynamic stochastic general equilibrium model (DSGE) of the type used, for example in Clarida, Galí, and Gertler (1999).

Consider an economy populated by a representative agent and a policy maker, and suppose there are two variables, y_t and x_t, being under the control of the agent and the policy maker, respectively. Suppose that y_t is determined according the following model:

$$y_t = \beta E_t y_{t+1} + \gamma x_t + u_t \tag{1}$$

where E_t denotes expectations conditional on information available at time t, and u_t is a shock, which is unobserved by the econometrician. The above equation can be thought of as a (possibly linearized version of) an Euler equation that determines the optimal choice of y_t by the agent given their objective function. The parameters β and γ will then be directly related to some 'deep' structural parameters that characterize the objective function (e.g., discount factors, elasticities, or price rigidity, depending on the meaning of y_t).

Suppose the policy variable x_t is determined according to the simple feedback rule:

$$x_t = \rho y_{t-1} + v_t \tag{2}$$

where v_t is an unobserved shock, which can be interpreted as a 'policy' shock and is assumed to be independent of the shock u_t. As an example, you can think of y_t as inflation and x_t as a measure of economic slack, e.g., unemployment or the output gap. In that case, equation (1) would be the new Keynesian Phillips curve and equation (2) could be thought of as some sort of backward-looking policy rule. The assumption that y_t enters equation (2) with a lag is only made to simplify the discussion of identification, since in that case we can justify treating x_t as exogenous in equation (1).

Now, consider the problem that an econometrician faces in trying to learn about the structural parameters β and γ from a sample of $\{y_t, x_t\}$. It is easy to see that without further restrictions, β and γ not separately identified. To see this, note that, under rational expectations, the law of motion (or reduced-form dynamics) of y_t is given by:[1]

$$y_t = \alpha x_t + \varepsilon_t, \quad \varepsilon_t = (1 - \alpha\beta\rho)^{-1} u_t. \tag{3}$$

Therefore, $E_t y_{t+1} = \alpha^2 \rho x_t + \alpha\rho\varepsilon_t$, and since ε_t is not observed by the econometrician, $E_t y_{t+1}$ needs to be instrumented by information that is known at time t and is orthogonal to u_t in equation (1). Under the stated assumptions, the optimal instrument is x_t, and $\hat{E}_t y_{t+1} = \alpha^2 \rho x_t$. But notice that $\hat{E}_t y_{t+1}$ is perfectly collinear with x_t, which is also included in the model as a regressor, so the rank condition for identification fails. There is no hope of recovering β by using further lags of x_t or y_t as instruments, since they are not relevant.

Next, consider what happens if policy changes over time, i.e., ρ in equation (2) is replaced by ρ_t. Because equation (1) is assumed to represent the behavior of agents, its parameters must be immune to the Lucas critique and therefore invariant to policy changes. However,

---

[1] This can be verified using the method of undetermined coefficients. In case of indeterminacy, equation (3) will be the minimum state variable solution

time variation in ρ will affect the reduced form for y_t, which now becomes

$$y_t = \alpha_t x_t + \varepsilon_t.$$

This implies the correlation between y_t and x_t changes over time. (Moreover, the reduced form innovations ε_t are heteroskedastic even when the structural shocks u_t and v_t are homoskedastic.) Returning to the issue of identification, the optimal instrument now becomes $\hat{E}_t y_{t+1} = \alpha_{t+1}\alpha_t \rho_t x_t$, and this is no longer collinear with the exogenous regressor x_t.[2] Thus, policy shifts induce exogenous variation that helps identify the parameters. This is true if there is at least one break in the parameter ρ, in which case the parameters β, γ will be just-identified. In other words, even though the number of instruments x_t is less than the number of parameters, i.e., identification seemingly fails on the order condition, a break in the parameters means we can find an additional relevant instrument and achieve identification.

Now, consider GMM estimation of the parameters of the model (1). The assumptions of the model imply the conditional moment restrictions $E\left[h_t(\theta)\,|x_t, y_{t-1}, x_{t-1}, ...\right] = 0$ where $\theta = (\beta, \gamma)'$ and $h_t(\theta) = y_t - \beta y_{t+1} - \gamma x_t$. By the reduced-form equation (3), the optimal instrument is x_t, so the above moment condition can be written as:

$$E\left[x_t h_t(\theta)\right] = 0, \quad \text{for all } t. \tag{4}$$

Let Z_t denote a vector of k instrumental variables that includes x_t as well as other predetermined variables, and let $Y_t = (y_{t+1}, x_t)$. The GMM moment vector is given by the k vector:

$$F_T(\vartheta) = \frac{1}{T}\sum_{t=1}^{T} Z_t\left(y_t - Y_t \vartheta\right) \tag{5}$$

where ϑ is some value of the parameters. Inference on θ is based on the distance of the sample moments $F_T(\vartheta)$ from zero. This corresponds to the *population* moment condition

$$E\left[F_T(\theta)\right] = 0 \tag{6}$$

which is noticeably weaker than the condition (4). Specifically, there are T restrictions in (4)

---

[2] It may help to think of ρ_t as predetermined at t, so that α_{t+1} is known in time t, but this is inessential for the argument.

but only $k \ll T$ in (6).

The assumptions of the model imply (4), but we GMM effectively only uses the weaker condition (6). This means that GMM reduces the $T \times 2$ dimensional information in the sample of $\{y_t, x_t\}$ to the statistic $T^{-1} \sum_{t=1}^{T} Z_t \left(y_t \vdots Y_t \right)$, which represents the average covariance between the instruments and the regressors over the full sample. Therefore, GMM ignores subsample variation in $\{y_t, x_t\}$ that leaves $T^{-1} \sum_{t=1}^{T} Z_t \left(y_t \vdots Y_t \right)$ constant. This can be justified only when the distribution of $\{y_t, x_t\}$ is strictly stationary, which implies, at a minimum, that the policy parameter ρ in equation (2) is constant. This is because any variation in $\{y_t, x_t\}$ over the sample that is orthogonal to $T^{-1} \sum_{t=1}^{T} Z_t \left(y_t \vdots Y_t \right)$ is uninformative about the parameters θ, i.e., $T^{-1} \sum_{t=1}^{T} Z_t \left(y_t \vdots Y_t \right)$ is sufficient for inference on θ. However, when ρ changes, the above discussion shows that this is no longer true, and hence ignoring the subsample variation in $\{y_t, x_t\}$ foregoes information about θ.

The main results of the paper can now be outlined in terms of the above example. First, we show in a formal statistical sense that unless we preclude *a priori* any breaks or time-variation in the reduced-form parameter ρ, inference based on the full-sample moment conditions (5) do not use all the relevant information. Second, we show how we can efficiently exploit the information in the model's assumptions (4) without making any assumptions about whether the model is identified, or whether there are any breaks in the reduced-form parameters, and without needing to know the number and timing of the breaks. In fact, our methods do not require that we specify any model for the reduced-form dynamics of the forcing variable x_t.

3 Theory

Consider a p-dimensional vector of structural parameters θ whose parameter region Θ is a subset of $\mathbb{R}^p$, and suppose that we observe a sample of size T given by a triangular array of random variables $\{Y_{t,T}\}_{t=1}^{T}$. For notational convenience, we will drop the dependence of the random variables in the sample on T. We assume that economic theory gives rise to a set of k moment conditions that can be represented in terms of a k-dimensional function of data and parameters $f_t(\theta)$, whose expectation vanishes at the true value of θ, i.e.,

$$E\left[f_t(\theta)\right] = 0 \quad \text{for all } t = 1, \ldots, T. \tag{7}$$

Our interest is in testing the null and alternative hypotheses:

$$H_0 : \theta = \theta_0 \quad \text{and} \quad H_1 : \theta \neq \theta_0, \tag{8}$$

using tests with significance level α. The robustness requirement is that α level tests should not reject H_0 more often than the nominal level asymptotically for a wide range of data generating processes (DGPs), under which a multivariate invariance principle applies to the sample moments, see Mueller (2008) for a motivation. There is a large class of tests that meet that requirement, and we shall therefore also address the question of efficiency by means of weighted average power (WAP) criteria.

Let

$$F_{sT}(\theta) = \sum_{t=1}^{[sT]} f_t(\theta) \tag{9}$$

where $[x]$ denotes the integer part of x and $s \in [0, 1]$. We refer to $F_{sT}(\theta)$ as partial-sample moment conditions, and $F_T(\theta)$ as full-sample moment conditions. Moreover, let $V_T(\theta)$ be an estimator of the variance of $T^{-1/2}F_T(\theta)$, and define for convenience the normalized vector

$$F^*_{sT}(\theta) \equiv V_T(\theta)^{-1/2} F_{sT}(\theta) \tag{10}$$

where $A^{-1/2}$ denotes the symmetric square root of a positive definite matrix A.

A typical Euler equation model with G equations gives rise to a set of conditional moment restrictions of the form $E[h_t(\theta) | \mathcal{I}_{t-1}] = 0$, where $h_t(\theta)$ is a G-dimensional function of data and parameters, e.g., a vector of residuals or structural errors, and $\mathcal{I}_{t-1}$ is the information set at time $t-1$. Given any set of instrumental variables $Z_t \in \Re^{G\times k}$ in $\mathcal{I}_{t-1}$, the conditional moment restrictions can be converted to unconditional restrictions in (7) by defining

$$f_t(\theta) = Z_t' h_t(\theta). \tag{11}$$

The single-equation linear instrumental variable (IV) model as well as the simultaneous equations model are special cases where $h_t(\theta)$ is linear.

We will not make any assumptions about the behavior of the Jacobian of the sample moment (when it exists), except when we test hypotheses on subsets of θ. Thus, we consider tests of H_0 that are only based on the partial-sample moments (9). The following condition suffices to obtain α-level tests of H_0.

Condition 1 *The normalized partial-sample moments $F^*_{sT}(\theta)$ defined in equation (10) satisfy*

$$T^{-1/2}\left\{F^*_{sT}(\theta_0) - E\left[F^*_{sT}(\theta_0)\right]\right\} \Rightarrow B_k(s), \quad s \in [0,1] \tag{12}$$

for all $\theta_0 \in \Theta$, where $B_k(s)$ is a k-dimensional standard Brownian motion.

Condition 1 is a high level assumption that the partial-sample moments satisfy a functional central limit theorem (FCLT). It can be derived from primitive conditions, e.g., assumptions 1, 7 and 8 in Sowell (1996). For instance, when the moment functions are given by equation (11), condition 1 will be satisfied when $h_t(\theta_0)$ is strong mixing with finite moments of order greater than 2, and Z_t is asymptotically mse stationary, see Hansen (2000). Asymptotic mse stationarity is weaker than strict stationarity and allows for non-permanent changes in the marginal distribution of Z_t.

Condition 1 is stronger than the assumption that the full-sample moments $T^{-1/2}F_T(\theta)$ satisfy a CLT, which suffices to obtain an identification-robust test of H_0 based on the continuously updated GMM objective function

$$\text{GMM-AR}(\theta_0) = \frac{F_T(\theta_0)' V_T(\theta_0)^{-1} F_T(\theta_0)}{T} = \frac{F^*_T(\theta_0)' F^*_T(\theta_0)}{T}. \tag{13}$$

This statistic was proposed by Stock and Wright (2000), and it is a generalization of the Anderson-Rubin test statistic in the linear IV regression model, see Stock, Wright, and Yogo (2002).

GMM-AR(θ_0) can be thought of as a statistic that tests the validity of the moment conditions, $E[f_t(\theta_0)] = 0$. In the case of equation (11), it is a Wald statistic for testing that Z_t can be excluded from the auxiliary regression of $h_t(\theta_0)$ on Z_t.[3] This test has no power against alternatives that leave the average moment conditions unchanged, namely $E[F_T(\theta_0)] = 0$. On the other hand, the necessary and sufficient condition for global identification of the parameters is $E[f_t(\theta_0)] \neq 0$ if $\theta \neq \theta_0$, for some $t = 1, \ldots, T$. So, tests based on the full-

---

[3] When $h_t(\theta_0)$ is G-dimensional, the auxiliary regression is a system of seemingly unrelated regressions (SUR).

sample GMM-AR statistic do not make use of all the restrictions that are implied by the original model given by equation (7).

For every θ, we can decompose the process $F^*_{sT}(\theta)$ into the random vector $F^*_T(\theta)$ and the process

$$F^B_{sT}(\theta) = F^*_{sT}(\theta) - sF^*_T(\theta). \tag{14}$$

A straightforward implication of condition 1 is that $F^*_T(\theta)$ and $F^B_{sT}(\theta)$ are asymptotically independent.

Lemma 2 *Under condition 1, the process $F^B_{sT}(\theta)$ defined in equation (14) satisfies*

$$T^{-1/2}\left\{F^B_{sT}(\theta_0) - E\left[F^B_{sT}(\theta_0)\right]\right\} \Rightarrow BB_k(s), \qquad s \in [0,1] \tag{15}$$

for all $\theta_0 \in \Theta$, where $BB_k(s)$ is a k-dimensional Brownian Bridge, which is independent of $B_k(1)$ defined in equation (12).

The statistic $F^B_{sT}(\theta)$ can be thought of as measuring variation in the data that is asymptotically orthogonal to the average moment conditions, and hence to the GMM-AR statistic (13). $F^B_{sT}(\theta)$ is asymptotically ancillary for θ if its limiting distribution does not depend on the true value of θ. By lemma 2, this depends on the limiting behavior of $T^{-1/2}E\left[F^B_{sT}(\theta_0)\right]$, or, equivalently, of $T^{-1/2}F^*_{sT}(\theta_0)$.

Let $m(\theta_0; t/T) = T^{\kappa}E\left[V^{-1/2}(\theta_0) f_t(\theta_0)\right]$, where κ distinguishes local ($\kappa = 1/2$) versus permanent ($\kappa = 0$) violations of the moment conditions under H_1. Assume $m(\theta_0; s)$ is bounded for all finite θ_0 and $s \in [0,1]$. It follows that $T^{\kappa-1}E\left[F^*_{sT}(\theta_0)\right] \to \int_0^s m(\theta_0; r)\,dr \equiv M(\theta_0; s)$, and $T^{\kappa-1}E\left[F^B_{sT}(\theta_0)\right] \to M(\theta_0; s) - sM(\theta_0; 1)$. Hence, $F^B_{sT}(\theta_0)$ is ancillary for θ if and only if $M(\theta_0; s) = sM(\theta_0, 1)$ for all $s \in [0,1]$ and $\theta_0 \in \Theta$, or $m(\theta_0; s)$ is independent of s.[4] This is equivalent to say that $E\left[f_t(\theta_0)\right]$ is (at least approximately) independent of t for all $\theta_0 \in \Theta$, that is $E\left[f_t(\theta_0)\right] = T^{-\kappa}m(\theta_0) + o\left(T^{-1/2}\right)$.

In fact, asymptotic linearity of the partial-sample moment conditions is implied by the usual regularity conditions for GMM, see Newey and McFadden (1994). However, such restrictions are typically both unnecessary to justify the validity of full-sample GMM, see e.g.,

---

[4] The result follows by differentiating both sides of equation $\int_0^s m(\theta_0, r)\,dr = sM(\theta_0; 1)$ with respect to s.

Li and Mueller (2006), as well as too strong in typical applications, as it will be demonstrated further below.

We have established that, in general, it pays to use $T^{-1/2}F_{sT}^{B}(\theta_0)$ in addition to $T^{-1/2}F_{T}^{*}(\theta_0)$ for inference. The next question is how to combine these two statistics efficiently. Since our objective is to do inference using weak assumptions, a suitable approach to asymptotic efficiency is the one proposed by Mueller (2008). This approach can be described briefly as follows.

Let $X_T(s)$, $s \in [0,1]$, denote a k-dimensional stochastic process that converges weakly to a process $X(s)$, continuous on $[0,1]$, and suppose we wish to test the hypothesis H_0 : $E[X(\cdot)] = 0$ against $H_1 : E[X(\cdot)] \neq 0$ using an asymptotically efficient test. Suppose that H_1 implies that $X(\cdot)$ is a draw from $C[0,1]^k$ using the measure μ_1 defined by the stochastic differential equation

$$dX(s) = dB(s) + \nu(s)\,ds \tag{16}$$

where $B(s)$ is a k dimensional standard Brownian motion and $\nu(s)$ is a nonstochastic continuous function, and H_0 implies the measure μ_0 defined by equation (16) with $\nu(s) = 0$. Thus, the null and alternative hypotheses can be equivalently stated as $H_0 : v(\cdot) = 0$ against $H_1 : \nu(\cdot) \neq 0$.

The optimal asymptotic level α test of H_0 against H_1 rejects for large values of the Radon-Nikodym derivative (Likelihood Ratio) of μ_1 with respect to μ_0, see Sowell (1996, Theorem 2), which is given by

$$\frac{d\mu_1}{d\mu_0}(X) = \exp\left\{\int_0^1 \nu(s)'\,dX(s) - \frac{1}{2}\int \nu(s)'\,\nu(s)\,ds\right\}. \tag{17}$$

Now, suppose $\nu(s)$ is a function of some parameters $\tau \in [0,1]^m$, with $\tau = (\tau_1, ..., \tau_m)'$ and $0 \leq \tau_1 \leq \ldots \leq \tau_m \leq 1$, and $\delta \in \Re^q$, that is, $\nu(s) = g(s|\tau, \delta)$. This setup allows for m deterministic shifts in the mean of X occurring at the fractions of the sample given by the vector τ. (We can equivalently think of $g(s|\tau, \delta)$ as stochastic, and consider the analysis as being conditional on the realization of the breaks. We comment on this alternative approach below.) The mean in each of the $m+1$ regimes is characterized by the parameter δ. In the leading case when $g(s|\tau, \delta)$ is linear in δ, the latter measures the mean in each regime, and we can write $g(s|\tau,\delta) = D(s|\tau)\,\delta$, where $D(s|\tau) \in \Re^{k\times q}$ is a matrix of indicator functions. So, the statistic in (17) specializes to

$$\xi(\tau,\delta) = \exp\left\{\delta' \int_0^1 D(s|\tau)'\,dX(s) - \frac{1}{2}\delta' \left[\int D(s|\tau)'\,D(s|\tau)\,ds\right]\delta\right\}$$

Under the above assumption, the alternative H_1 is composite, and efficient tests can be obtained by WAP criteria. Notice that H_0 can be equivalently stated as $\delta = 0$, and hence τ is a nuisance parameter that is only present under H_1. Hence, the testing problem belongs to the class of nonregular problems studied by Andrews and Ploberger (1994), where it is well-known that optimal WAP tests generally depend on the choice of weights over the possible alternatives.

Let $P(\tau)$ denote a distribution of weights over the possible break dates, and let $P(\delta|\tau)$ denote the distribution of weights over the range of δ conditional on τ. Then, the optimal test rejects for large values of the statistic $\int\int \xi(\tau,\delta)\,dP(\delta|\tau)\,dP(\tau)$. Since the distribution of that statistic does not depend on τ or δ under H_0, critical values can be obtained easily by Monte Carlo simulation for any choice of weight functions $P(\tau)$ and $P(\delta|\tau)$.

Following Andrews (1993), Andrews and Ploberger (1994) and Sowell (1996), suppose that $P(\delta|\tau)$ is normal with zero mean and variance $c\Sigma(\tau)$, where $\Sigma(\tau)$ is a positive definite matrix and c is a scalar that controls the weight attached to different alternatives for δ. Smaller (larger) values of c put less (more) weight on alternatives that are further away from the null of $\delta = 0$. This approach yields tests that maximize power against alternatives that are 'equally hard to detect'. Let $Z(\tau) = \int_0^1 D(s|\tau)'\,dX(s)$, and $\Sigma(\tau) = \left[\int D(s|\tau)'\,D(s|\tau)\,ds\right]^{-1}$. Then, the optimal test statistic becomes

$$\int \exp\left\{\frac{1}{2}\frac{c}{1+c} Z(\tau)'\,Z(\tau)\right\} dP(\tau)$$

and it depends both on the weights over τ as well as on the value of c. This is a special case of the test derived by Andrews and Ploberger (1994), when there are no parameters to be estimated under the null. The familiar supremum, exponential and average functionals of Andrews and Ploberger (1994) are derived by taking limits as $c/1+c$, c and $1/c$ go to infinity, respectively.

We now apply the above approach to the problem of testing (8). An important difficulty in this case is that local power analysis also requires assumptions about identification of the model, i.e., about the rank of the Jacobian of the moment conditions in a neighborhood of θ_0, which are difficult to justify in practice. To be more specific, let $J(\theta_0; s) = \partial M(\theta_0; s)/\partial\theta'$ and suppose $E\left[T^{-1}\partial F_{sT}(\theta_0)/\partial\theta'\right] \Rightarrow J(\theta_0; s)$. Then, local identification requires that the $k \times p$ Jacobian matrix $J(\theta_0; s)$ is of full rank at the true value $\theta = \theta_0$ for some $s \in [0, 1]$. In the canonical constant parameter case, $J(\theta_0; s)$ is linear in s, so this is equivalent to the usual full rank condition for the limiting Jacobian of the average moment conditions $T^{-1}F_T(\theta_0)$. More generally, the rank condition is equivalent to either (i) full-rank value of the full-sample Jacobian, $J(\theta_0; 1)$; or (ii) full rank of the Jacobian of $T^{-1}F_{sT}^{BB}(\theta_0)$, $J(\theta_0; s) - sJ(\theta_0; 1)$. The latter requires that the model is not contiguous to a model with a constant Jacobian, i.e., that there are *permanent* breaks in the Jacobian. It is clear that both (i) and (ii) can be too strong in practice, see, e.g., Stock, Wright, and Yogo (2002) and Li and Mueller (2006). So, we shall propose methods of inference that do not require any of the above two assumptions.

Optimality results under weak identification exist only for the case of the linear IV regression model with a single endogenous regressor, see Andrews, Moreira, and Stock (2006), and there is currently no extension of those results even for linear GMM with constant parameters. Given this difficulty, we take an alternative approach. Instead of testing the hypotheses in (8) directly, we look for optimal tests of the validity of the model's identifying restrictions under H_0, namely:

$$H_0^* : E\left[f_t(\theta_0)\right] = 0, \quad \text{against} \quad H_1^* : E\left[f_t(\theta_0)\right] \neq 0$$

where $E\left[f_t(\theta_0)\right]$ may be time-varying under H_1^*.

To apply the aforementioned optimality theory, we need to make some assumptions about the behavior of the statistic $F_{sT}^*(\theta)$ under the alternative hypothesis H_1^*. Note that these assumptions are only used to motivate a particular choice of test statistic for exploiting the stability restrictions amongst a large class of similar tests. Specifically, we assume that under H_1^*, $E\left[f_t(\theta_0)\right] = D(s|\tau)\,\delta$, with $m = 1$, i.e., we allow for only two regimes (one break) under the alternative. The parameter δ measures the violation of the moment conditions in each of the two regimes under the alternative

Optimal tests of H_0^* can then be obtained by the following split-sample moment conditions, which are obtained by splitting the sample at any point $s \in (0,1)$:

$$\widetilde{F}_T(\theta; s) = \begin{pmatrix} F_{sT}(\theta), \\ F_T(\theta) - F_{sT}(\theta) \end{pmatrix} \quad \text{and} \quad \widetilde{V}_T(\theta; s) = \begin{pmatrix} V_{1T}(\theta; s) & 0 \\ 0 & V_{2T}(\theta; s) \end{pmatrix}$$

Note that for the estimators of $V_{iT}(\theta; s)$ of the long-run variance in each sample, we could use the respective partial-sample estimators, or simply the full-sample estimator $V_T(\theta)$. Both are acceptable under condition 1.

Next, define the 'split-sample' GMM objective function as:

$$S_T(\theta_0; s) = \frac{1}{T} \widetilde{F}_T(\theta_0; s)' \, \widetilde{V}_T(\theta_0; s)^{-1} \, \widetilde{F}_T(\theta_0; s), \quad s \in (0,1). \tag{18}$$

Observe that $S_T(\theta_0; s)$ is approximately distributed as $\chi^2(2k)$ for any given s under condition 1. An asymptotically α-level test of the hypothesis $H_0 : \theta = \theta_0$can be obtained by comparing the statistic $S_T(\theta_0; s)$ to the $1-\alpha$ quantile of the $\chi^2(2k)$ distribution. Moreover, under 2, the statistic

$$AR_T^b(\theta_0; s) = S_T(\theta_0; s) - \text{GMM-AR}(\theta_0)$$

is asymptotically independent of GMM-AR(θ_0). Intuitively, $AR_T^b(\theta_0; s)$ is asymptotically equivalent to a quadratic form of $F_{sT}^b(\theta_0)$, and it is therefore capture information that is orthogonal to the full-sample moments.

Optimal WAP tests are obtained by applying the approach of Andrews and Ploberger (1994). With equal weights over alternatives violations of the moment conditions that are 'equally hard' to detect, the optimal test statistic takes the form

$$\text{exp-S}_T(\theta_0) \equiv (1+c)^{-k} \int_{\mathcal{S}} \exp\left\{ \frac{1}{2} \frac{c}{1+c} S_T(\theta_0, s) \right\} dP(s).$$

The following statistic can be equivalently written in terms of GMM-AR(θ_0) and $AR_T^b(\theta_0; s)$, and corresponds to putting equal weight to violations of the average moment conditions versus instability in the moment conditions under H_1. By putting different weights on those alternatives, we can derive alternative tests. For instance, the standard GMM-AR statistic of

Stock and Wright (2000) can be obtained by putting zero weight on instability of the moment conditions under H_1. The opposite case is to place zero weight on average violations of the moment conditions. The resulting test takes the form

$$\text{exp-AR}_T^{BB}\left(\theta_0\right) \equiv \left(1+c\right)^{-k/2} \int_{\mathcal{S}} \exp\left\{\frac{1}{2}\frac{c}{1+c} AR_T^b\left(\theta_0, s\right)\right\} dP\left(s\right)$$

and can be thought of as a pure stability test for $E\left[f_t\left(\theta_0\right)\right]$, conditional on $E\left[F_T\left(\theta_0\right)\right]=0$.

We shall consider the three forms of the above statistics, familiar from Andrews and Ploberger (1994).

$$\begin{aligned}
\text{ave-AR}\left(\theta_0\right) &\equiv \int_{\mathcal{S}} S_T\left(\theta, s\right) dP\left(s\right) \\
\text{exp-AR}\left(\theta_0\right) &\equiv \log \int_{\mathcal{S}} \exp\left[\frac{1}{2} S_T\left(\theta, s\right)\right] dP\left(s\right) \\
\text{sup-AR}\left(\theta_0\right) &\equiv \sup_{s\in\mathcal{S}} S_T\left(\theta, s\right).
\end{aligned}$$

The first and second functionals are special cases of the functional proposed by Andrews and Ploberger (1994, Eq. 1.1), with $c=0$ and ∞, respectively. The third is a quasi likelihood ratio test of the validity of the moment conditions. The key difference with the usual stability test statistics is that we evaluate the base statistic ($Q_T\left(\theta_0; s\right)$ here) under $H_0 : \theta = \theta_0$, rather than at the GMM estimator of θ obtained by minimizing the full-sample objective function (13). The asymptotic distributions of these statistics under the null hypothesis can be obtained, under Condition 1 using the continuous mapping theorem.

Theorem 3 *Under condition 1 and under the null hypothesis* $H_0 : \theta = \theta_0$,

$$\begin{aligned}
\textit{ave-}AR\left(\theta_0\right) &\Rightarrow \int_{\mathcal{S}} \left\{\frac{B_k\left(s\right)' B_k\left(s\right)}{s} + \frac{\left[B_k\left(1\right)-B_k\left(s\right)\right]' \left[B_k\left(1\right)-B_k\left(s\right)\right]}{1-s}\right\} dP\left(s\right) \\
\textit{exp-}AR\left(\theta_0\right) &\Rightarrow \log \int_{\mathcal{S}} \exp\left\{\frac{B_k\left(s\right)' B_k\left(s\right)}{2s} + \frac{\left[B_k\left(1\right)-B_k\left(s\right)\right]' \left[B_k\left(1\right)-B_k\left(s\right)\right]}{2\left(1-s\right)}\right\} dP\left(s\right) \\
\textit{sup-}AR\left(\theta_0\right) &\Rightarrow \sup_{s\in\mathcal{S}} \left\{\frac{B_k\left(s\right)' B_k\left(s\right)}{s} + \frac{\left[B_k\left(1\right)-B_k\left(s\right)\right]' \left[B_k\left(1\right)-B_k\left(s\right)\right]}{1-s}\right\}.
\end{aligned}$$

Critical values for the above limiting distributions can be easily obtained by simulation.

4 Simulations

We assess the finite-sample size and power of the above statistics and compare them with their full-sample counterpart, the GMM-AR statistic, using Monte Carlo simulations. We start by examining a prototypical linear instrumental variables (IV) regression model, because in this model we have full control over the amount of information in the sample through the so-called concentration parameter, see Stock, Wright, and Yogo (2002).

4.1 Linear IV regression

The model consists of a structural and a reduced-form equation. The structural equation is

$$y_{1,t} = \theta y_{2,t} + u_t, \quad t = 1, ..., T \tag{19}$$

where $\{y_{1,t}, y_{2,t}\}_{t=1}^{T}$ is a sequence of observed variables, $\{u_t\}_{t=1}^{T}$ is an unobserved error, and $\theta \in \Re$ is the unknown structural parameter. The reduced-form equation (also known as first-stage regression) is given by

$$y_{2,t} = Z_t \pi_t + v_{2,t}, \quad t = 1, ..., T \tag{20}$$

where $Z_t \in \Re^{1\times k}$, $t = 1, ..., T$ is a sequence of observed instrumental variables that are fixed (i.e., nonstochastic), $\{v_{2,t}\}$ is an unobserved (reduced-form) error, and $\pi_t \in \Re^{k\times 1}$, $t = 1, ..., T$ is a sequence of unknown parameters. The errors u_t and $v_{2,t}$ are i.i.d. and have a mean zero bivariate normal distribution. The identifying restrictions in this model are $E(Z_t u_t) = 0$ for all t, and the moment function $f_t(\theta)$ in equation (7) is $f_t(\theta) = Z_t (y_{1,t} - \theta y_{2,t})$.

The constant-parameter IV regression model is a special case of this setup with $\pi_t = \pi$ for all t. The assumption of normality of the errors implies the availability of a low-dimensional sufficient statistic, which in the constant parameter case is given by $\sum_{t=1}^{T} Z_t' Y_t \in \Re^{k\times 2}$, see Andrews, Moreira, and Stock (2006). The linear IV model is a curved exponential model, since the dimension of the minimal sufficient statistic is higher than the number of parameters when $k > 1$, and this has implications for inference, see Garderen (1997).

When $\pi_t \neq \pi$, the aforementioned statistic is no longer sufficient. From the factorization theorem, a sufficient statistic is given by the sequence $\{Z_t' Y_t\}_{t=1}^{T}$. Since the dimension of the sufficient statistics is much higher than $k \times 2$, we see that time variation in the parameters increases the curvature of the model.

It is well-known that in the constant-parameter IV regression model, the amount of information in the data about the structural parameters (or the quality of instruments/identification) can be characterized using a unitless measure known as the 'concentration parameter', which is $\lambda = \sum_{t=1}^{T} \pi' Z_t' Z_t \pi$. We can think of the contribution of each observation to the identification of θ as being equal to $\pi' Z_t' Z_t \pi$, but when π_t is time-varying, the incremental information is $\pi_t' Z_t' Z_t \pi_t$, and so the total amount of information is $\lambda = \sum_{t=1}^{T} \pi_t' Z_t' Z_t \pi_t$.

The DGP is given by equations (19) and (20). We assume that $E\left(u_t^2\right) = E\left(v_{2,t}^2\right) = 1$ and $E\left(u_t v_{2,t}\right) = \rho_{uv}$, and we consider the cases $\rho_{uv} = \{0, 0.5, 0.95\}$ (results for $\rho_{uv} = 0.95$ are omitted because they were very similar). There is a single instrument Z_t ($k = 1$), drawn independently from a $N(0, 1)$ distribution and kept fixed in repeated samples. In all of the simulations, we set $T = 300$. The number of Monte Carlo replications used is 1000 for power curves and 2000 for the size of tests. The random numbers used in all experiments are the same across different values of the parameters. All stability tests use uniform weights over the range $s \in [0.15, 0.85]$. We consider m breaks in the first-stage regression that define the $m+1$ regimes/subsamples $\frac{j-1}{m+1}T < t \leq \frac{j}{m+1}T$ for $j = 1, .., m+1$. We use an odd number of breaks and set the coefficients in each regime such that they rise linearly from $-\Pi$ (first regime) to Π (last regime), so that $\Pi_0 = 0$ and there is no identification through full-sample statistics. This DGP can also be written in constant parameter form by defining the $(m+1)$-dimensional vector $\widetilde{Z}_t$ taking the value $Z_t e_j$, where e_j is the unit vector with 1 in position j, in each regime $j = 1, ..., m+1$. Hence, the total number of relevant instruments is $m+1$, and the model is overidentified (when time variation in Π is taken into account). We report power curves for $\lambda = \{3, 10\}$, corresponding to weak and strong identification, respectively.

Figure 1 compares the power curves of the usual full-sample S statistic, which is the Anderson-Rubin (AR) statistic in this case, with the power of the three stability test statistics defined above. We also report the power of two other test statistics that are based on all the relevant instruments $\widetilde{Z}_t$ and require knowledge of the exact break dates. The first is the AR statistic, and the second is Moreira's (2003) conditional Likelihood ratio (CLR) statistic, which has some optimality properties (within a class of tests), see Andrews, Moreira, and

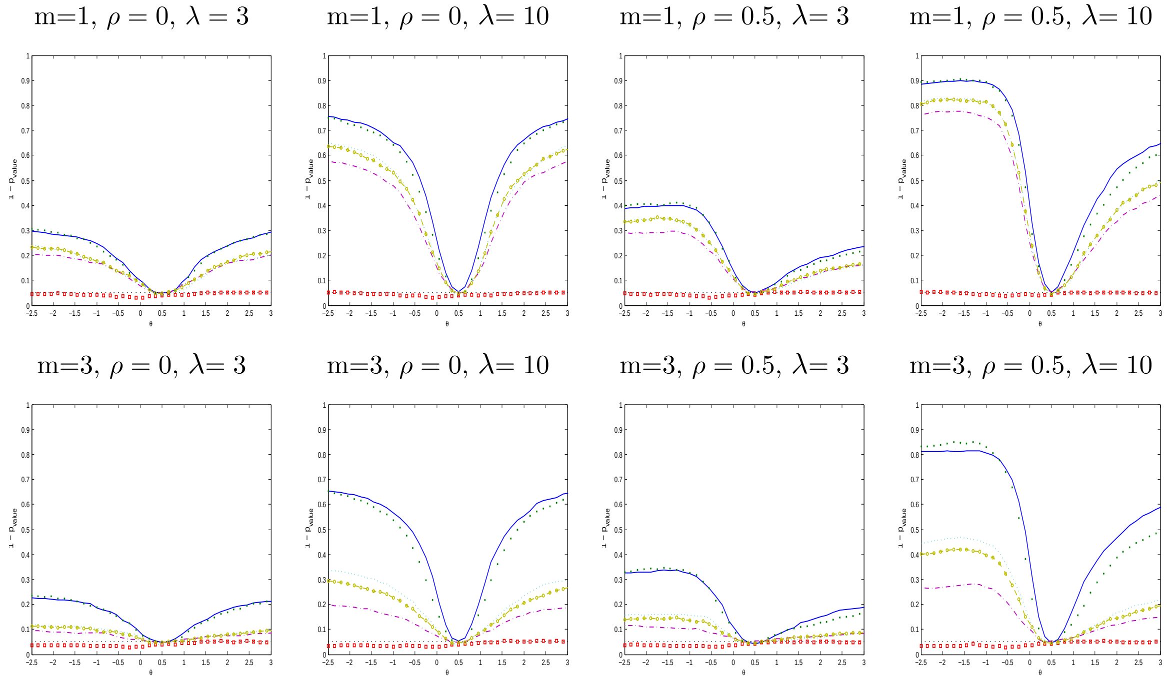
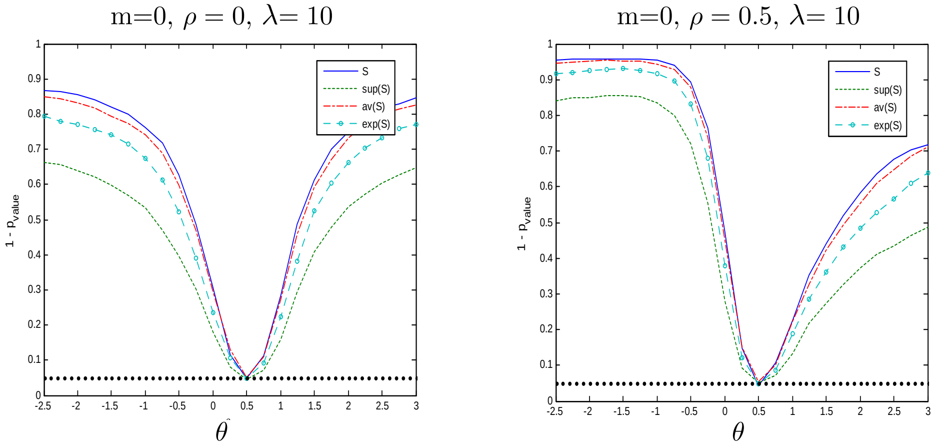


Figure 1: Power curves of 5% level tests of the null hypothesis $H_0 : \theta = 0.5$ with $T = 300$ computed using 1000 MC replications. m is the number of breaks in the first-stage regression coefficient, ρ is the correlation between structural and reduced form errors, μ is the concentration parameter. The curves correspond to: split-sample CLR (solid), split-sample AR (crosses), ave-AR (dash-dotted), exp-AR (dash-circle), sup-AR (dotted), full-sample AR (squares)

Stock (2006). These two statistics are infeasible in practice, but we report them to get a sense of how much power is lost by using statistics that do not require knowledge of the break date.

Three main findings of these results are noteworthy. First, the full-sample AR test, which would be optimal if there were no instabilities, has no (nontrivial) power in all cases, as expected, since $\Pi_0 = 0$. Second, all three stability tests have considerable power even when there are multiple breaks. Third, when there is only one break stability tests are dominated by the infeasible AR and CLR tests but this is not the case when there are multiple breaks. This shows that we can go a long way in extracting the information in the instability of Π without knowing the number/magnitude/timing of breaks.

Next, we ask the exact opposite question: how much do we lose by using stability tests when, in fact, Π is constant? The DGP is the same as in the previous design but with $m = 0$, $\Pi_t = \Pi$ for all t, and $\lambda = 10$ (strong identification). In this case the AR test is the optimal



m=0, $\rho = 0$, λ= 10

m=0, $\rho = 0.5$, λ= 10

Figure 2: Power comparisons when there are no instabilities in the DGP

test, since the model is just-identified. The results, reported in Figure 2, suggest that the stability tests are not as powerful as the AR test, as expected, but they still have considerable power. In fact, the power curve of the ave-AR test is very close to that of the AR test. The combination of these results with those of of the previous experiment suggest that use of the stability tests offers considerable power advantages when there are instabilities in the reduced form, with little cost when the reduced form is stable.

4.2 A forward-looking model

Consider the following data generating process given by the equations:

$$y_t = \beta E_t(y_{t+1}) + \gamma x_t + \varepsilon_t \tag{21}$$

$$x_t = \rho_{1,t} x_{t-1} + \rho_{2,t} x_{t-2} + v_t. \tag{22}$$

If y_t denotes inflation, and x_t a measure of marginal costs, then equation (21) can be thought of as the new Keynesian Phillips curve (NKPC). The reduced form model for the forcing variable given by equation (22) does not have a structural interpretation, but it represents a good description of the reduced form dynamics in actual data, see, e.g., Mavroeidis (2005).The structural parameters of interest are $\theta = (\beta.\gamma)$, and the moment conditions are given by the equation (11) with $h_t(\theta) = y_t - \beta y_{t+1} - \gamma y_t$ and Z_t denotes predetermined variables.

The above model differs from the linear IV regression (19) in two important ways. First, the residuals of the model are autocorrelated by construction and they need not be homoskedastic (economic theory does not preclude time-variation in volatilities, for example). Second, the instruments are typically lags of the regressors, which are predetermined but not strongly exogenous, so any instability in the marginal distribution of the data affects the distribution of the instruments, as well.

Solving equation (21) forward, we find that $y_t = \alpha_t x_t + \alpha_t \rho_{2,t} \beta x_{t-1} + \varepsilon_t$, where $\alpha_t = \gamma \,/ \left[1 - \beta\left(\rho_{1,t} + \beta\rho_{2,t}\right)\right]$, so that the reduced-form model for y_t and x_t is a restricted (time-varying parameter) vector autoregression

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \underbrace{\begin{pmatrix} \phi_{1,t} & \phi_{2,t} \\ \rho_{1,t} & \rho_{2,t} \end{pmatrix}}_{\Pi_t} \underbrace{\begin{pmatrix} x_{t-1} \\ x_{t-2} \end{pmatrix}}_{Z_t} + \begin{pmatrix} \eta_t \\ v_t \end{pmatrix}.$$

It is straightforward to show that, when $\rho_{1,t}$ and $\rho_{2,t}$ are constant, so is Π_t and its determinant is proportional to $\left(\alpha\rho_2\right)^2$. Hence, since α is proportional to γ, the rank condition for identification is satisfied if and only if $\gamma \neq 0$ and $\rho_2 \neq 0$, see Kleibergen and Mavroeidis (2009). However, when $\rho_{1,t}, \rho_{2,t}$ are time-varying, this condition is no longer necessary for identification.

We simulate the null rejection frequencies of the GMM-AR statistic and the stability statistics ave-AR, exp-AR and sup-AR with uniform weights over the range $[0.15, 0.85]$. The null hypothesis is $H_0 : \beta = 0, \gamma = 0.1$. The reduced-form errors are drawn from a Normal distribution with $\sigma_\eta^2 = 1, \sigma_{\eta v} = 0.1$ and $\sigma_v^2 = 1$, and the autoregressive coefficients are set according to $\rho_{1,t} = \rho\left(1 - \rho_{2,t}\right)$, with $\rho = 0.9$. These parameters were calibrated to US data on inflation and the labor share, and are taken from Kleibergen and Mavroeidis (2009). We use two lags of y_t and x_t as instruments, i.e., $k = 4$. We consider three sample sizes $T = \{100, 200, 500\}$ and use 10,000 Monte Carlo replications. We also consider the two alternative estimators of the Weight matrix for the stability statistics based on partial samples and the full sample, respectively.

The parameter $\rho_{2,0}$ is calibrated to 0.05 over the full-sample 1960-2008q3. We examine two types of breaks in the parameters. The first type is a single break in the middle of the sample. The estimated value of ρ_2 is -0.1 over the first subsample and the change in ρ_2 is estimated at 0.3 with a standard error of 0.15 (so a Chow test at that break date would just reject the null of stability at the 5% level). We consider both the calibrated value as well as two other situations, with break magnitudes that are two and four standard deviations larger than in the data, respectively, i.e., $\rho_{2,t} = \{-.2, .4\}$ and $\{-.3, .6\}$.

The second type of break is continuous time-variation in the parameters. In this case,

$$\rho_{2,t} = \rho_{2,0} + \frac{\kappa}{T^{1/2}} B\left(\frac{t}{T}\right)$$

where $B(t/T)$ is a scalar Brownian motion process. This case is referred to as martingale time variation (MTV). The draws of $\rho_{2,t}$ are kept fixed in repeated samples, so the rejection frequencies are conditional on $\{\rho_{2,t}\}$.

The results are reported in Table 1 and we can draw the following conclusions. First, the stability tests never overreject, though some of them seem to be under-sized. Second, the rejection frequency of the tests is insensitive to the magnitude of the break in the reduced form parameters. This is especially interesting in the cases of a single-break in $\rho_{2,t}$, denoted SB, because they are not covered by condition 1.[5] Nonetheless, we do not see any size distortions in those cases, and the rejection frequencies are not sensitivity to the magnitude of the break as T increases.

5 Empirical applications

We consider two prominent Euler equation models that form the core of the canonical new Keynesian Policy model which plays a central role in macroeconomic policy debate. Both equation can be written as special cases of the equation:

$$y_t = \beta E_t y_{t+1} + \sum_{j=1}^{p} \gamma_j y_{t-j} + \lambda x_t + \varepsilon_t. \tag{23}$$

The first model is the new Keynesian Phillips curve (NKPC) with indexation, studied by Sbordone (2005) and Christiano, Eichenbaum, and Evans (2005), where y_t is inflation,

---

[5] This is because permanent shifts in the autoregressive coefficients of x_t change the variance of both x_t and y_t, and therefore also change the variance of the instruments, which are lags of (y_t, x_t)..

		GMM-AR		ave-AR		exp-AR		sup-AR	
		10%	5%	10%	5%	10%	5%	10%	5%
$T = 100$	no break	9.06	4.25	8.16	3.34	6.97	2.63	4.10	1.45
	SB: 0.3	9.34	4.21	8.2	3.38	6.96	2.75	4.09	1.57
	SB: 0.6	9.31	4.26	8.16	3.41	6.9	2.77	4.09	1.54
	SB: 0.9	9.28	4.42	8.21	3.44	6.96	2.76	4.29	1.61
	MTV: $\kappa = 1$	9.61	4.53	7.79	3.54	6.48	2.77	3.97	1.37
	MTV: $\kappa = 3$	9.52	4.53	7.71	3.47	6.46	2.72	4.03	1.40
	MTV: $\kappa = 5$	9.46	4.45	7.80	3.52	6.71	2.79	3.99	1.52
	MTV: $\kappa = 9$	9.36	4.56	7.99	3.63	6.86	2.92	4.18	1.52
$T = 200$	no break	9.91	4.80	9.02	4.37	7.89	3.72	5.79	2.54
	SB: 0.3	10.01	4.78	8.93	4.49	7.90	3.72	5.73	2.57
	SB: 0.6	10.03	4.79	9.01	4.58	8.05	3.82	5.91	2.70
	SB: 0.9	9.95	4.90	9.00	4.45	7.94	3.86	5.88	2.73
	MTV: $\kappa = 1$	9.83	4.59	8.97	4.17	7.97	3.77	6.13	2.59
	MTV: $\kappa = 3$	9.72	4.62	9.01	4.18	7.91	3.92	6.14	2.57
	MTV: $\kappa = 5$	9.84	4.62	9.07	4.14	7.93	3.86	6.22	2.54
	MTV: $\kappa = 9$	9.64	4.7	8.97	4.12	7.95	3.82	6.19	2.64
$T = 500$	no break	9.65	4.56	9.33	4.24	9.15	4.38	7.73	3.67
	SB: 0.3	9.59	4.66	9.29	4.30	8.76	4.34	7.85	3.68
	SB: 0.6	9.73	4.64	9.53	4.56	8.83	4.24	7.95	3.68
	SB: 0.9	9.74	4.69	9.49	4.54	9.22	4.29	7.90	3.71
	MTV: $\kappa = 1$	9.85	4.68	9.65	4.67	9.22	4.34	7.47	3.80
	MTV: $\kappa = 3$	9.86	4.57	9.53	4.73	9.25	4.31	7.5	3.75
	MTV: $\kappa = 5$	9.83	4.57	9.5	4.68	9.11	4.33	7.56	3.76
	MTV: $\kappa = 9$	9.90	4.62	9.57	4.75	9.13	4.34	7.53	3.76

Table 1: Rejection frequencies of GMM-AR, ave-AR, exp-AR and sup-AR tests under the null hypothesis $\beta = 0, \gamma = 0.1$ in the model $y_t = \beta E_t(y_{t+1}) + \gamma x_t + u_t$, $x_t = \rho_{1,t} x_{t-1} + \rho_{2,t} x_{t-2} + v_t$, with $\rho_{1,t} = 0.9(1 - \rho_{2,t})$. SB stands for single break at t=T/2, and number indicates $\Delta\rho_{2,T/2}$: value calibrated from data is 0.3 (st. error 0.15). MTV denotes martingale time variation with variance $\kappa T^{-1/2}$. Instruments include two lags of y_t and x_t.

$\beta = \delta / (1 + \delta \varrho)$, δ is a discount factor, ϱ is the fraction of prices that are indexed to past inflation, when not optimally set, α is the probability that a price will be fixed in a given period, $\gamma_1 = 1/(1 + \delta\varrho)$, $\gamma_j = 0$ for $j > 1$ and $\lambda = (1 - \alpha)(1 - \delta\alpha) / (\alpha (1 + \delta\varrho))$. The variable x_t is a measure of economic slack or marginal costs. In our empirical analysis, we investigate specifications with alternative measures of the forcing variable x_t: the labor share, used in Galí and Gertler (1999) and Sbordone (2002), the CBO measure of the output gap, and the unemployment rate. Inflation is measured by the GDP deflator. We use three lags of y_t and x_t as instruments, and we also consider the implications of adding the Federal Funds rate to the instrument set. The parameter δ is fixed to 1, in accordance with the literature, see Kleibergen and Mavroeidis (2009) and we allow for an unrestricted constant in the equation.

The second model is the Euler equation for output, see Fuhrer and Rudebusch (2004), where y_t represents the output gap, and x_t denotes the ex ante real interest rate, $\lambda = -\sigma$ and σ denotes the intertemporal elasticity of substitution. In the canonical version of the model, see Woodford (2003), $\beta = 1$ and $\gamma_j = 0$. In the unrestricted version, the lags of y_t can be thought of as representing adaptive expectations, and $\beta \in [0, 1]$ represents the fraction of rational agents.

5.1 Euler equation for output

We compute joint 90% and 95% confidence regions for the coefficient β and the elasticity of substitution σ in the Euler equation for output (23). We consider the case $p = 2$ and use two lags of inflation, the output gap and the federal funds rate as instruments (results for the case $p = 3$ are very similar). The ave- exp- and sup-AR statistics are computed over the middle 70% of the sample, as before. The sample period is period 1966q1-2008q3, to match Fuhrer and Rudebusch (2004).

The results are reported in Figures 3, 4, and the following conclusions can be drawn. First, the 90%-level GMM-AR confidence interval for the coefficient β covers the entire parameter space, so this parameter is completely unidentified by information over the full sample. This conclusion remains robust to changes in the sample and number of instruments. Second, the confidence regions based on inverting the stability tests are a fraction of their full-sample counterparts. When we apply them to the full sample period 1966-2008, the confidence intervals are sufficiently tight to provide accurate inference about both β and σ. Third, when we look at the post-1983 period, the stability confidence intervals are much wider, though still smaller than their GMM-AR counterparts. This is consistent with the evidence of policy regime shifts in the first part of the sample, and relative stability in the post-1983 sample. Thus, our results suggest that instability in monetary policy over the late 1970s and early 1980s may have contributed to improving the identifiability of the structural parameters of the Euler equation for output.

5.2 NKPC

We compute joint 90% and 95% confidence regions for the coefficients α (price rigidity) and ϱ (indexation parameter) in the NKPC with alternative forcing variables. The ave- exp- and sup-AR statistics are computed over the middle 70% of the sample. The sample period is 1960q1-2008q3.

The results for the labor share as forcing variable are given in figure 5. We notice that the stability confidence regions are smaller than the GMM-AR regions. However, the difference is not very large, and the indexation coefficient remains very weakly identified. This suggests that the informational content of the stability restrictions is small for the NKPC. Figure 6 shows how the results change when we use the Federal Funds rate as an additional instrument. We notice that, even though the GMM-AR confidence region barely changes, the stability regions shrink. This is consistent with the view that information in the stability restrictions comes primarily from policy changes.

The results for the model that uses the output gap as the forcing variable are given in Figures 7 and 8. All confidence sets are tighter than for the case of the labor share, but the conclusions are exactly analogous to previous case. The main difference is that the price stickiness parameter is estimated to be much higher, indicating perhaps unreasonably long price durations (over 10 quarters).

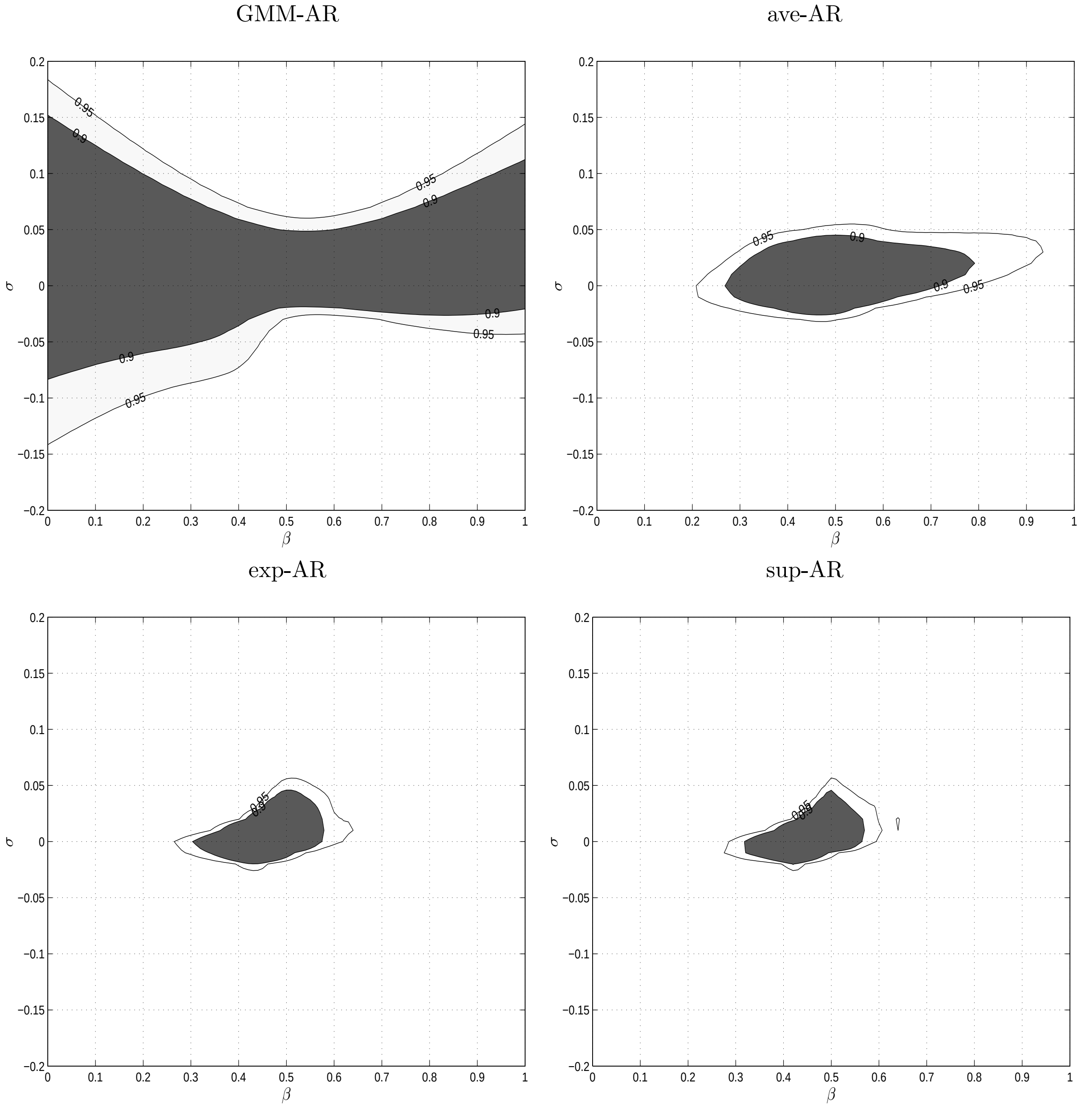


Figure 3: GMM-AR and stability-based confidence regions for β and σ in the Euler equation for output. Period: 1966q1 2008q3. Instruments include two lags of the output gap, inflation and the fed funds rate.

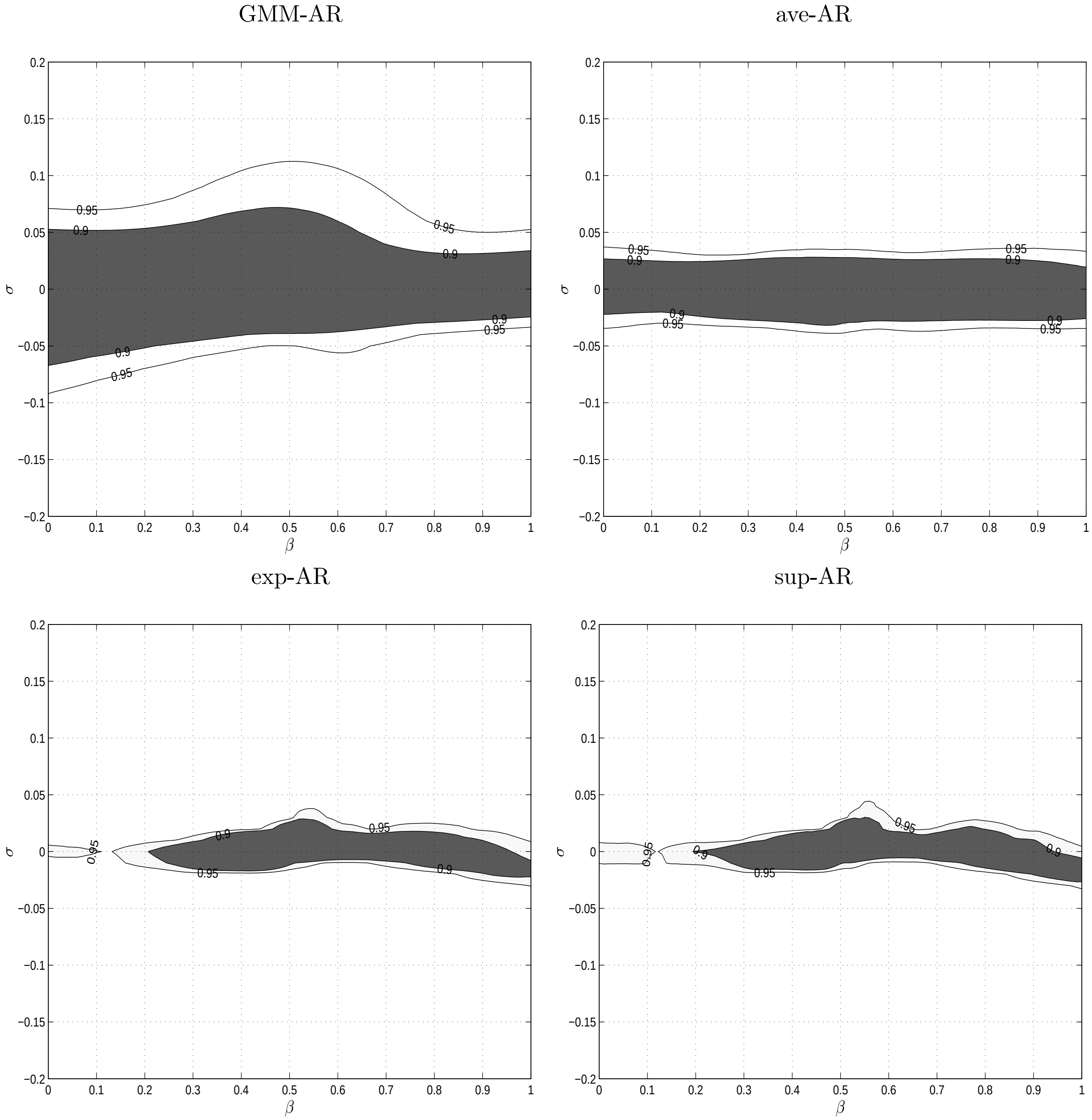


Figure 4: GMM-AR and stability-based confidence regions for β and σ in the Euler equation for output. Period: 1984q1 2008q3. Instruments include two lags of the output gap, inflation and the fed funds rate.

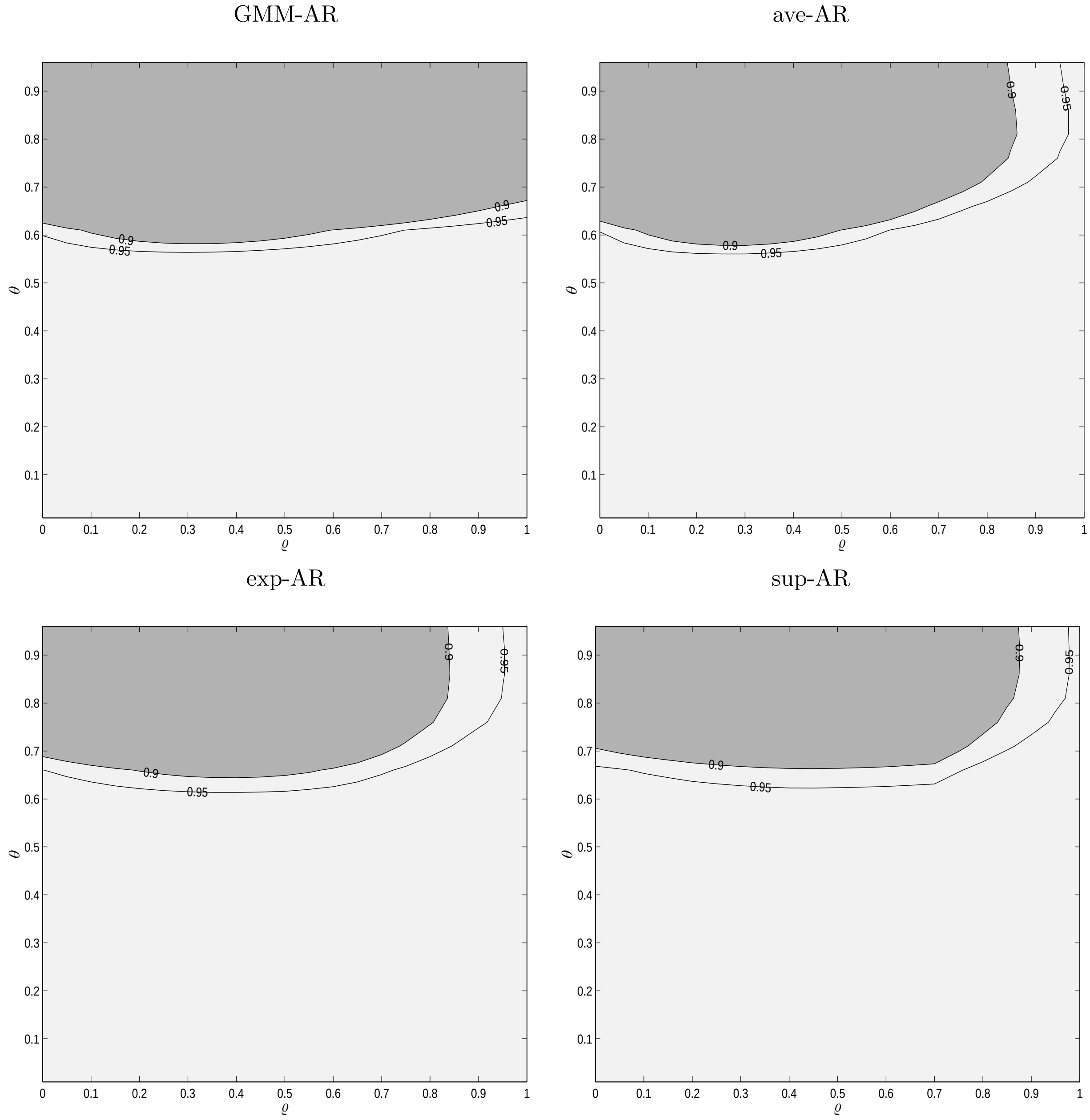


Figure 5: GMM-AR and the three stability-based confidence regions for α and ϱ in the NKPC. Period: 1960q1 2008q3. The forcing variable is the labor share. Instruments include three lags of inflation and the labor share.

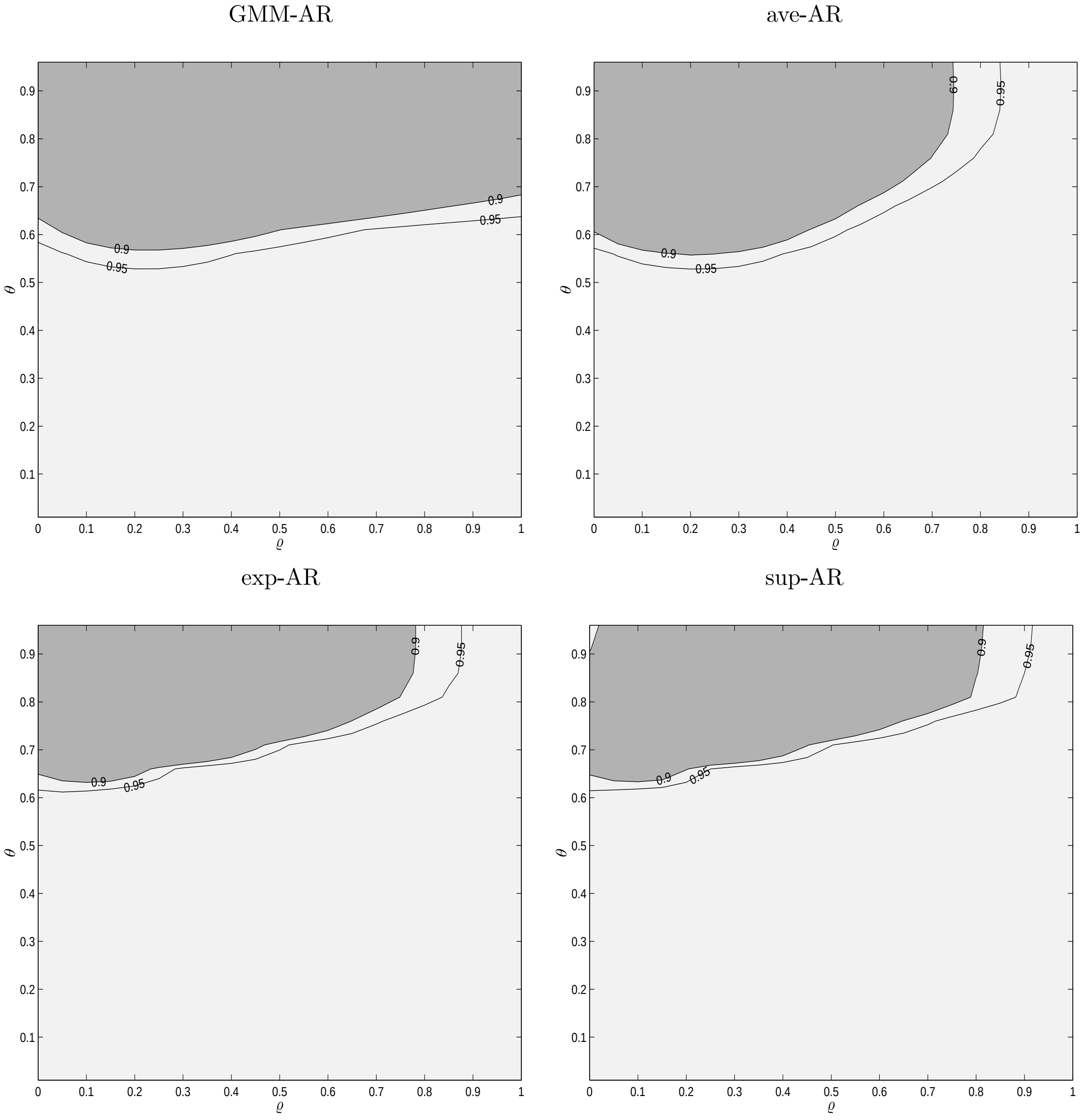
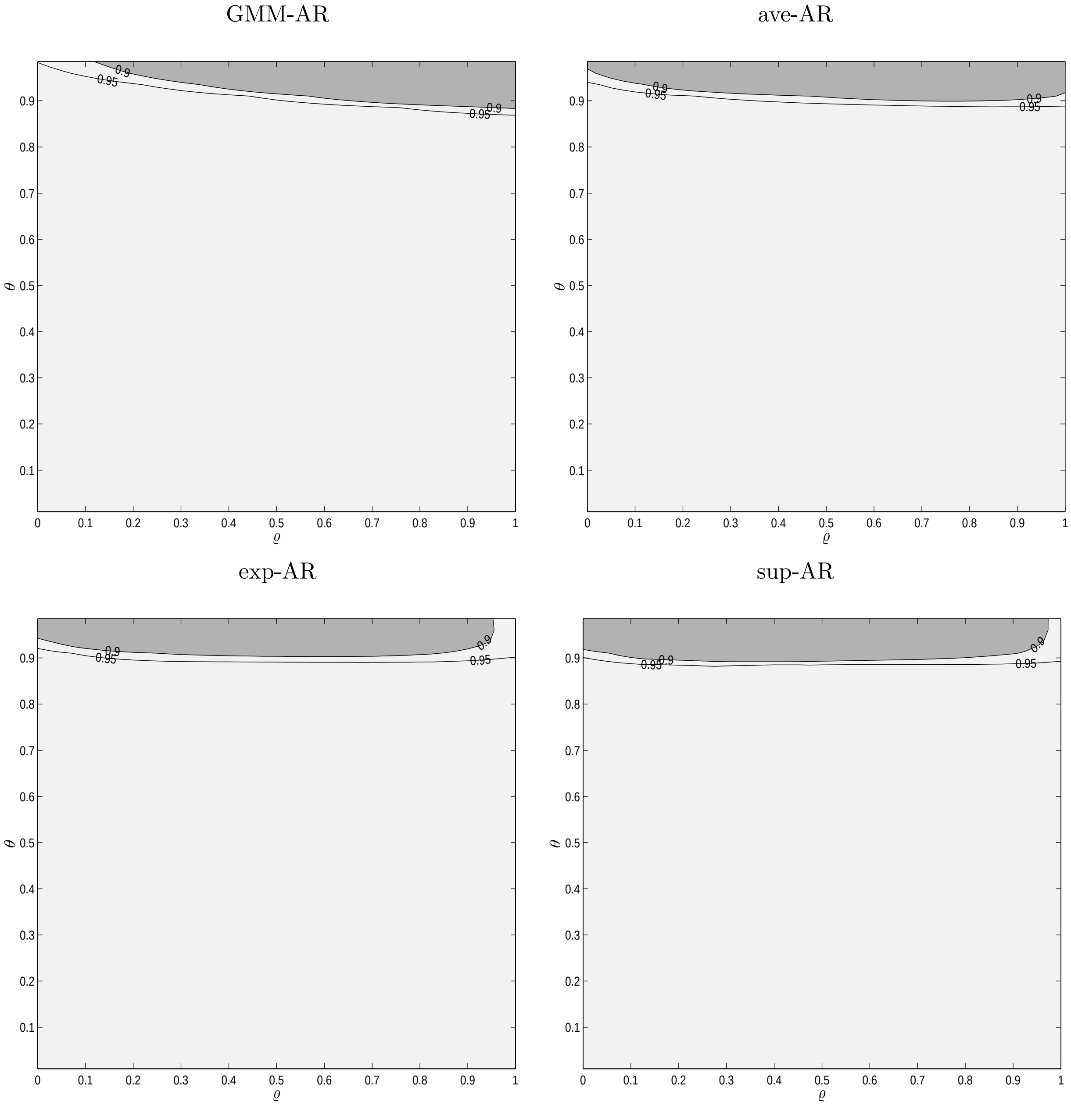


Figure 6: GMM-AR and the three stability-based confidence regions for α and ϱ in the NKPC. Period: 1960q1 2008q3. The forcing variable is the labor share. Instruments include three lags of inflation and the labor share and the Federal Funds rate.



GMM-AR

ave-AR

exp-AR

sup-AR

Figure 7: GMM-AR and the three stability-based confidence regions for α and ϱ in the NKPC. Period: 1960q1 2008q3. The forcing variable is the output gap. Instruments include three lags of inflation and the output gap.

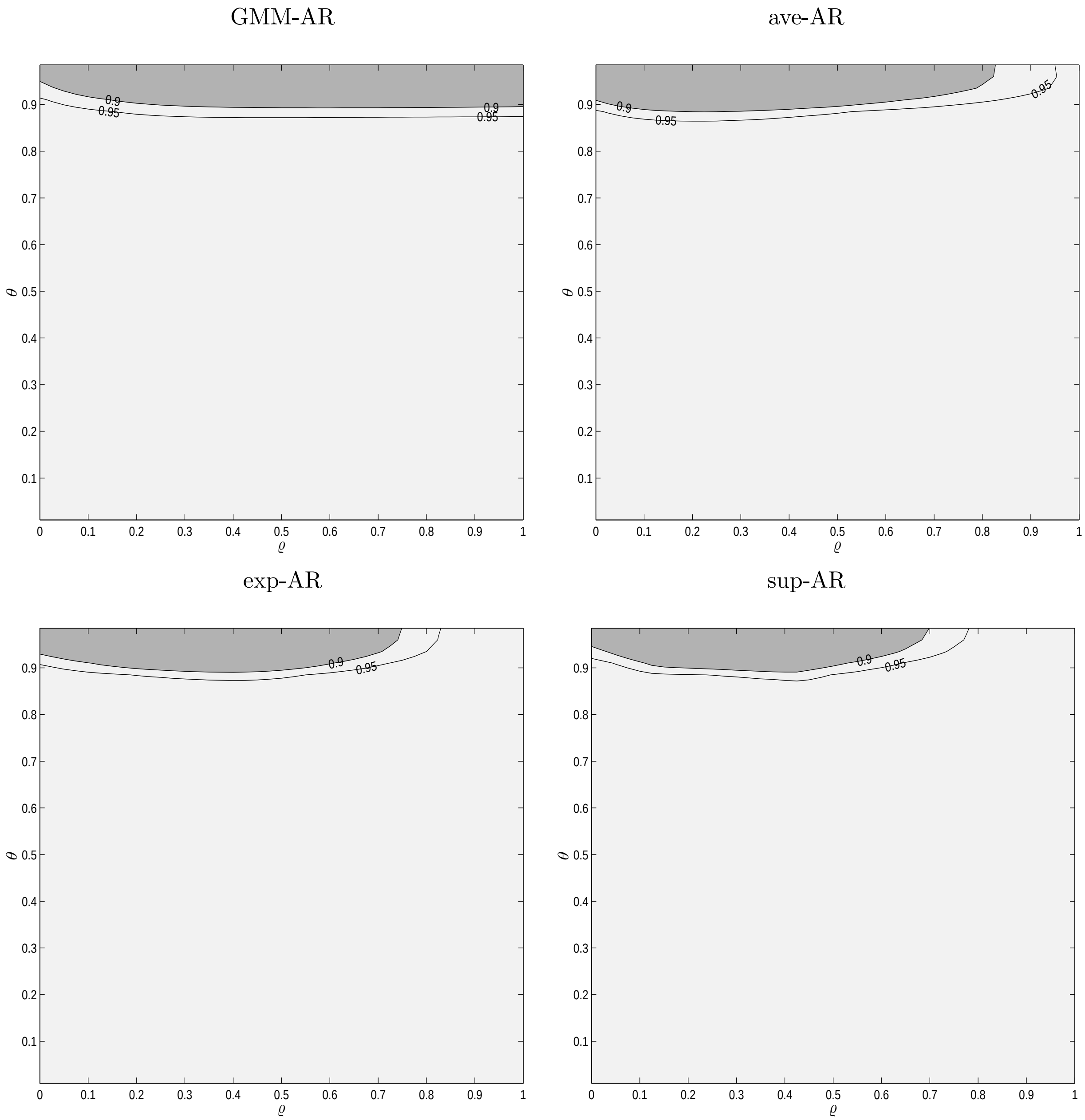


Figure 8: GMM-AR and the three stability-based confidence regions for α and ϱ in the NKPC. Period: 1960q1 2008q3. The forcing variable is the ouptup. Instruments include three lags of inflation, the output gap and the Federal Funds rate.

The results for the unemployment as the forcing variable are similar to the output gap case, with the main difference being that the stability confidence intervals for the indexation parameter are much smaller.

Finally, concerns over potential instabilities in the first part of the sample have led many researchers to estimate the NKPC over the period after the Volcker disinflation, typically after 1983. To investigate the robustness of the results over subsamples, we estimate the confidence regions over subsamples. The results are summarized in Table 2 in the form of 90% confidence intervals for each parameter. We notice that in most cases, the stability confidence intervals are smaller than GMM-AR confidence intervals. Two features are noteworthy. First, some of the confidence intervals in the pre-1984 sample are empty, indicating that the NKPC may be subject to the Lucas critique over that period. Similar results were reported in Kleibergen and Mavroeidis (2009). Second, the confidence intervals vary similarly over the two periods across all specifications: prices seem to have become stickier after 1984, i.e., the Phillips curve may have become flatter, while indexation seems to be less important now than it was before 1984.

6 Conclusions

In this paper we addressed the following issues. First, we showed that instability of the reduced form parameters is highly informative for inference on structural parameters that are stable, i.e., immune to the Lucas critique. We demonstrated this both in theory as well as in practice. Secondly, we proposed methods for extracting the information in the stability restrictions.

An interesting feature of our proposed method of inference is that it allows for identification of the parameters even when the usual GMM order condition for identification is not satisfied, i.e., when the number of instruments is smaller than the number of parameters. This may be useful in situations where alternative exclusion restrictions may be controversial.

Forcing variable	Sample period	Statistic	90% confidence intervals	
			Price stickiness (α)	Indexation (ϱ)
Labor Share				
	1960q1-1983q4	GMM-AR	[0.41, 1.00]	[0.00, 1.00]]
		Ave-AR	[0.56, 1.00]	[0.05, 1.00]
		Exp-AR	$\emptyset$	$\emptyset$
		Sup-AR	$\emptyset$	$\emptyset$
	1984q1-2008q2	GMM-AR	[0.61, 1.00]	[0.00, 0.60]
		Ave-AR	[0.56, 1.00]	[0.00, 0.70]
		Exp-AR	[0.61, 1.00]	[0.00, 0.60]
		Sup-AR	[0.61, 1.00]	[0.00, 0.60]
Output gap				
	1960q1-1983q4	GMM-AR	[0.86, 1.00]	[0.15, 1.00]
		Ave-AR	[0.91, 1.00]	[0.30, 1.00]
		Exp-AR	$\emptyset$	$\emptyset$
		Sup-AR	$\emptyset$	$\emptyset$
	1984q1-2008q2	GMM-AR	[0.00, 0.56]$\cup$[0.76, 1.00]	[0.00, 1.00]
		Ave-AR	[0.86,1.00]	[0.00, 0.53]
		Exp-AR	[0.86, 1.00]	[0.00, 0.45]
		Sup-AR	[0.84, 1.00]	[0.00, 0.43]
Unemployment				
	1960q1-1983q4	GMM-AR	[0.76, 1.00]	[0.00,1.00]
		Ave-AR	[0.84,1.00]	[0.00,1.00]
		Exp-AR	[0.94, 1.00]	[0.65, 1.00]
		Sup-AR	$\emptyset$	$\emptyset$
	1984q1-2008q2	GMM-AR	[0.84, 1.00]	[0.00, 0.45]
		Ave-AR	[0.89, 1.00]	[0.00, 0.30]
		Exp-AR	$\emptyset$	$\emptyset$
		Sup-AR	$\emptyset$	$\emptyset$

Table 2: 90% confidence intervals on the coefficients of the NKPC. Instruments include three lags of inflation and forcing variable.

References

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica 61*(4), 821–56.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2006). Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica 74*(3), 715–752.

Andrews, D. W. K. and W. Ploberger (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica 62*(6), 1383–1414.

Christiano, L. J., M. Eichenbaum, and C. Evans (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *J. Political Economy 113*, 1–45.

Clarida, R., J. Galí, and M. Gertler (1999). The Science of Monetary Policy: A New Keynesian Perspective. *Journal of Economic Literature 37*(4), 1661–1707.

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.

Elliott, G. and U. K. Mueller (2006). Efficient tests for general persistent time variation in regression coefficients. *Review of Economic Studies 73*(4), 907–940.

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

Galí, J. and M. Gertler (1999). Inflation dynamics: a structural econometric analysis. *Journal of Monetary Economics 44*, 195–222.

Garderen, K. J. v. (1997). Curved exponential models in econometrics. *Econometric Theory 13*(6), 771–790.

Hansen, B. E. (2000). Testing for structural change in conditional models. *Journal of Econometrics 97*(1), 93–115.

Kleibergen, F. (2005). Testing parameters in GMM without assuming that they are identified. *Econometrica 73*(4), 1103–1123.

Kleibergen, F. and S. Mavroeidis (2009). Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve. *Journal of Business and Economic Statistics*. forthcoming.

Li, H. and U. Mueller (2006). Valid inference in partially unstable gmm models. mimeo, Princeton University.

Lucas, R. E. J. (1976). Econometric policy evaluation: a critique. In K. Brunner and A. Meltzer (Eds.), *The Philips Curve and Labor Markets.*, Carnegie-Rochester Conference Series on Public Policy. Amsterdam: North-Holland.

Mavroeidis, S. (2005). Identification issues in forward-looking models estimated by GMM with an application to the Phillips Curve. *Journal of Money Credit and Banking 37*(3), 421–449.

Mueller, U. (2008). An alternative sense of asymptotic efficiency. mimeo.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden (Eds.), *The Handbook of Econometrics*, Volume 4, pp. 2111–2245. North-Holland.

Newey, W. K. and K. D. West (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Perron, P. (2005). Dealing with structural breaks. Boston University - Department of Economics - Working Papers Series WP2005-017, Boston University - Department of Economics.

Sbordone, A. M. (2002). Prices and unit labor costs: a new test of price stickiness. *Journal of Monetary Economics 49*, 265–292.

Sbordone, A. M. (2005). Do expected future marginal costs drive inflation dynamics? *Journal of Monetary Economics 52*(6), 1183–1197.

Sims, C. A. and T. Zha (2006). Were There Regime Switches in U.S. Monetary Policy? *American Economic Review 96*(1), 54–81.

Sowell, F. (1996). Optimal tests for parameter instability in the generalized method of moments framework. *Econometrica 64*(5), 1085–1107.

Stock, J. H. and M. W. Watson (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics 14*(1), 11–30.

Stock, J. H. and J. H. Wright (2000). GMM with weak identification. *Econometrica 68*(5), 1055–1096.

Stock, J. H., J. H. Wright, and M. Yogo (2002). GMM, weak instruments, and weak identification. *Journal of Business and Economic Statistics 20*, 518–530.

Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton University Press.