UvA Econometrics

Discussion Paper: 2004/06

# Weak identification of forward-looking models in monetairy economics

Sophocles Mavroeidis

**www.fee.uva.nl/ke/UvA-Econometrics**

Department of Quantitative Economics
Faculty of Economics and Econometrics
Universiteit van Amsterdam
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA Universiteit van Amsterdam

# Weak identification of forward-looking models in monetary economics*

SOPHOCLES MAVROEIDIS

*Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands (e-mail: s.mavroeidis@uva.nl)*

## Abstract

Recently, single-equation GMM methods have become popular in the monetary economics literature, for estimating forward-looking models with rational expectations. We discuss a method for analyzing the empirical identification of such models that exploits their dynamic structure and the assumption of rational expectations. This allows us to judge the reliability of the resulting GMM estimation and inference and reveals the potential sources of weak identification. With reference to the New Keynesian Phillips curve of Galí and Gertler (1999) and the forward-looking Taylor rules of Clarida, Galí, and Gertler (2000), we demonstrate that the usual 'weak instruments' problem can arise naturally, when the predictable variation in inflation is small relative to unpredictable future shocks (news). Hence, we conclude that those models are less reliably estimated over periods when inflation has been under effective policy control.

*JEL classification:* C22, E31.
*Keywords:* New Phillips Curve, Taylor Rule, concentration parameter, Rational Expectations, GMM

## I Introduction

Forward-looking models are commonly used in monetary economics both by academics and practitioners, in order to advise on, or assess the efficacy of, monetary policy. In recent years, small-scale forward-looking macro models have been increasingly used by central banks around the world to examine broader issues of monetary policy, especially so relative to the traditional large-scale macro models of the seventies.

*I would like to thank Adrian Pagan for very helpful comments.

There are both theoretical and practical reasons for this growing popularity. On theoretical grounds, first, by explicitly incorporating forward-looking components, these models address the Lucas (1976) critique, which reduced-form models do not. Second, because they are usually built on micro-foundations, it is argued that they represent underlying economic structure. Moreover, their so-called 'structural' parameters admit interesting economic interpretations, and thus they are more appealing than the reduced-form models. Third, these models are based on 'rational expectations', which have become an essential feature of most macroeconomic models.

There is a widespread view that authorities, as well as economic agents, are forward looking in their behaviour. For instance, the monetary authority needs to look forward due to lags in the transmission mechanism that mean monetary policy takes time to have an effect, as seen by the following quote by a prominent central banker:

> 'The challenge of monetary policy is to interpret current data on the economy [...] with an eye to anticipating future inflationary forces and to countering them by taking action in advance.' (Alan Greenspan, Chairman of the Federal Reserve Board in his Humphrey-Hawkins testimony in 1994, cited in Batini and Haldane 1999, p. 157.

This prompted researchers to develop models of the form

$$\begin{array}{lll} \text{Pure:} & y_t & = \beta\, \mathrm{E}(y_{t+1}|\mathcal{F}_t) + e_t \\ \text{Hybrid:} & y_t & = \beta\, \mathrm{E}(y_{t+1}|\mathcal{F}_t) + \gamma\, y_{t-1} + e_t \end{array} \tag{1}$$

The former is a pure forward-looking model, whereas the latter is a hybrid version containing both forward and backward-looking adjustment. These models have been used to address the following questions that are central to the current monetary policy debate: (i) Are agents forward-looking? (Are expectations rational?) (ii) How important is forward-looking behaviour compared to 'backwardness'?

Two common applications of forward-looking models are found in monetary economics. One comes from the recent literature on monetary policy rules, where it has become common practice to estimate Taylor-type rules from historical data, see Taylor (1999) and the papers therein. One approach, popularized by Clarida, Galí, and Gertler (1998) and Clarida, Galí, and Gertler (2000), is the estimation of the reaction function parameters from a single equation of the form:

$$r_t = \bar{r} + \beta\,(\mathrm{E}(\pi_{t+j}|\mathcal{F}_t) - \bar{\pi}) + \gamma\,\mathrm{E}(x_{t+i}|\mathcal{F}_t) + \epsilon_t \tag{2}$$

where $r_t$, $\pi_t$ and $x_t$ denote the policy rate, inflation and output gap respectively, $\bar{r}$ and $\bar{\pi}$ denote the equilibrium rate and inflation target respectively, and $\mathrm{E}(\cdot|\mathcal{F}_t)$ denotes expectations conditional on the available information, and $i, j$ are specified.

Another important example is the influential paper of Galí and Gertler (1999), which uses the same econometric methodology in estimating the 'New Phillips curve', a forward-looking model for inflation dynamics:

$$\pi_t = \lambda s_t + \beta \mathrm{E}(\pi_{t+1}|\mathcal{F}_t) + \gamma \pi_{t-1} \tag{3}$$

where $s_t$ is the labour share. Other examples of forward-looking Phillips curves include the models proposed in Buiter and Jewitt (1989), Fuhrer and Moore (1995), Batini, Jackson, and Nickell (2000) and Galí, Gertler, and López-Salido (2001).

In view of the fact that such equations involve unobservable expectations of variables, researchers proceed as follows. They replace expectations by actual realizations of the variables and derive orthogonality conditions that may be used to estimate the parameters of the model with the Generalized Method of Moments (GMM). These moment conditions are derived based on the assumption of rational expectations, i.e., that the expectation-induced 'errors in variables' must be orthogonal to the information set available to the agents, denoted $\mathcal{F}_t$ at the time the expectations are formed. The nature of the moment conditions guides the choice of the appropriate weighting matrix for the GMM estimator, i.e., what type of corrections should be made for serial correlation or heteroscedasticity of the residuals.

This approach is popular because it is relatively easy to implement. It apparently obviates the need to model the whole system of variables involved in the analysis, and in particular those that are thought of as 'exogenous'; it is known to be robust to a wider range of Data Generating Processes than FIML estimators (Hansen (1982)); and in general, it is expected to work well for the estimation of various types of Euler Equation models under weak conditions.

The robustness of this method arises not only with respect to specification errors in other equations of the system that one is not estimating. It also is also robust to another type of mis-specification. Full information methods require that the RE system be solved to derive the (restricted) reduced form or 'observable structure'. This observable structure, and hence the resulting likelihood, depends on whether the system has a 'forward' or 'backward' solution, which cannot be determined a priori, except by assumption, see Pesaran (1987, chapter 5). An advantage of limited information methods is that they do not require the solution of the model prior to estimation.

However, it is easy to see why such an approach invites criticism. First, it is not grounded on prior testing for the lack of feedbacks in the variables. This is a necessary condition for the absence of information loss in the estimation and inference on the parameters of interest. In fact, the properties of the non-modelled variables are crucial for the identification of the model's parameters, even when the former are thought to be 'exogenously' determined.

Secondly, pathological cases such as 'weak instruments', which are common across the spectrum of applied econometrics, are empirically relevant for those models, and have been shown to impart serious distortions on the distributions

of the estimators and test statistics, thus invalidating conventional inference, see e.g. Hansen, Heaton, and Yaron (1996) and other papers in that issue of the *Journal of Business and Economic Statistics*.

In this paper, we discuss a method for analyzing the identifiability of those models, based on a combination of the relevant economic and statistical theory. By economic theory we mean the application of 'rational expectations' to derive the reduced form of the system of all endogenous and exogenous variables. Then, the statistical theory provides us with a measure of the 'strength' of identification, which can be readily derived from that reduced form. This is known as the concentration parameter, measuring the predictability of the (future) endogenous regressors on current information *relative* to the genuinely unpredictable innovations.

The structure of the paper is as follows. Section II reviews the relevant theory of weak instruments. Section III discusses the weak identification for the New Keynesian Phillips curve and forward-looking monetary policy rules. Finally, section IV concludes. Algebraic derivations are given in the appendix.

# II Weak identification

The devastating implications of weak identification for GMM estimation and inference have been well-documented in a growing theoretical and applied literature in the 1990s, see Stock, Wright, and Yogo (2002) for a review. The important lesson from that literature is that the usual rank condition for identification of structural models (e.g. simultaneous equations or IV regressions) is not sufficient to guarantee reliable inference using GMM *in finite samples*. How informative any given sample is for the parameters of interest can be judged by the expected bias and size distortions of conventional GMM estimators and test statistics.

Traditionally, large distortions have been attributed to problems of 'small samples'. However, the weak instruments literature has shown that such distortions are not necessarily a small sample problem. Rather, they depend on the amount of information that is present in the data for the parameters of interest. As shown by Stock, Wright, and Yogo (2002), this information is to some (large) extent characterized by the so-called concentration parameter, which will be introduced below. This is a unitless measure of the 'quality' of the instruments, akin to a signal-noise ratio in the first-stage regression of the endogenous variables on the instruments.

Before proceeding, it is important to define the terms partial or under-identification, weak identification and weak instruments. Consider a parametric model specified in terms of a set of orthogonality conditions. The true value of the parameters is defined as the point in the parameter space where the orthogonality conditions vanish. A parameter is unidentified on a given information set if the resulting orthogonality conditions vanish for more than one value of this parameter. The structural model is *partially* or *under*-identified

if any function or subset of its parameters is un-identified.

To distinguish between the general case in which the rank condition for identification is satisfied, and the more specific case when GMM is reliable we will adopt the terminology of Johansen and Juselius (1994, p.10). We refer to the former as *generic* identification. This includes both the case when GMM is reliable, which will be referred to as *empirical* or 'strong' identification, and the case when GMM is not reliable, which is commonly called weak identification.[1]

In linear models estimated using instrumental variables, weak identification is known as the 'weak instruments' problem, see Stock, Wright, and Yogo (2002), Stock and Wright (2000) and Wright (2003). Stock, Wright, and Yogo (2002) use the more general term 'weak identification' to describe weak instruments problems in the context of nonlinear GMM estimation or when the errors are heteroscedastic and/or serially correlated. Since forward-looking models estimated by GMM involve at least serially correlated errors, we will use the term weak identification in accordance with the above convention.

To illustrate the main implications of weak identification, we offer a simple exposition of this issue in the context of a univariate linear IV regression with fixed instruments. In this case, the analytics are simple and provide a useful insight into the more general asymptotic theory given in the literature, as well as a benchmark for interpreting the results of Monte Carlo experiments on the finite sample properties of GMM estimators.

## A primer on weak instruments

Consider the IV estimator of a parameter $\theta$ in the model (4):

$$y = Y\,\theta + u \tag{4}$$
$$Y = Z\,\Pi + v \tag{5}$$

where $(y, Y)$ is a $T \times (1+p)$ matrix of endogenous variables, $Z$ is a non-stochastic $(T \times k)$ matrix of instrumental variables, such that $\mathrm{E}[Z_t u_t] = 0$, $\lim_{T\to\infty} T^{-1}Z'Z = \Sigma_{ZZ}$, with $\mathrm{rank}(\Sigma_{ZZ}) = \mathrm{rank}(Z'Z) = k$ for all $T$, and $U = (u, v) \sim N(0,\ \Sigma_{UU} \otimes I_T)$. The quantity $\lambda = \Sigma_{vv}^{-1}\Sigma_{vu}$ measures the 'endogeneity' of $Y$ and determines the bias of the OLS estimator of $\theta$. It is more common to characterize this endogeneity by the correlation coefficient between $u$ and $v$, namely $\rho = \Sigma_{vv}^{1/2}\lambda\,\sigma_u^{-1}$.

The IV estimator of $\theta$ is:

$$\widehat{\theta}_{IV} = \left(Y'Z\,(Z'Z)^{-1}\,Z'\,Y\right)^{-1} Y'\,Z\,(Z'\,Z)^{-1}\,Z'\,y = \left(\widehat{\Pi}'\,(Z'\,Z)\,\widehat{\Pi}\right)^{-1}\widehat{\Pi}'Z'\,y. \tag{6}$$

where $\widehat{\Pi}$ is the OLS estimator of $\Pi$ in the 'first-stage' regression (5). When $\mathrm{rank}(\Pi) = p$, the limiting distribution of $\widehat{\theta}_{IV}$ follows from standard asymptotic

[1] The concept of weak identification is not specific to GMM. In likelihood inference, it refers to a situation in which the expected value of the likelihood function is flat around the true parameter, i.e. the information matrix at the true parameter is near singular.

theory:

$$\sqrt{T}\left(\widehat{\theta}-\theta_0\right)=\left(\widehat{\Pi}'\left(\frac{Z\,Z}{T}\right)\widehat{\Pi}\right)^{-1}\frac{1}{\sqrt{T}}\widehat{\Pi}'Z'u\xrightarrow{D}\mathsf{N}\left[0,\sigma_u^2\left(\Pi'\Sigma_{ZZ}\Pi\right)^{-1}\right]. \quad (7)$$

However, when $\text{rank}(\Pi) < p$, this approximation breaks down, see Phillips (1989). Moreover, the approximation becomes unreliable when $\Pi$ is 'close' to being of reduced rank, in a sense that will be made precise below.

**Lack of identification** It is easier to see what happens first in the univariate un-identified case with one instrument, where $p = k = 1$, with $\Pi = 0$. Defining $e_t = (u_t - v_t'\,\lambda)$, such that $\mathrm{E}(e_t v_t) = 0$, with variance $\sigma_{u.v}^2 = \sigma_u^2(1-\rho^2)$, the IV estimator can be written as:

$$\widehat{\theta}_{IV}=\theta_0+\frac{Z'u}{Z'v}=\theta_0+\lambda+\frac{Z'e}{Z'v}\sim(\theta_0+\lambda)+\frac{\sigma_{u.v}}{\sigma_v}\,t_1$$

where $t_1$ follows a Student's t-distribution with 1 degree of freedom (also known as the Cauchy distribution) and $\sigma_v^2 = \Sigma_{vv}$ has been introduced for notational simplicity in this special case.[2] This distributional result holds approximately (for large $T$), but also exactly (for any $T$), under the normality assumption for $(u, v)$ and the non-randomness of the instruments. Thus, we see that in the un-identified case, the IV estimator is far from normal and exhibits a 'double' inconsistency: it is $O_p(1)$ (i.e., its variability does not fall with $T$), and centered on the plim of the OLS estimator, which is $(\theta_0 + \lambda)$.

Next, we look at what happens when we add more irrelevant instruments, i.e., $k > 1$ and still $\Pi = 0$. This time, the distribution of the IV estimator becomes:

$$\widehat{\theta}_{IV}-\theta_0=\lambda+\frac{v'\,P_Z\,e}{v'\,P_Z\,v}\sim\lambda+\frac{\sigma_{u.v}}{\sigma_v\sqrt{k}}\,t_k$$

where $P_Z = Z\,(Z'Z)^{-1}Z'$ and $t_k$ is distributed as Student's t with $k$ degrees of freedom. We notice that the IV estimator now has moments up to the degree of over-identification, $k - 1$, and that its variance is falling linearly with the number of instruments.

**Weak identification** Building on the above discussion, we wish to investigate what happens when identification is 'weak', i.e., $\Pi \neq 0$ but 'close' to zero. One approach is to develop higher order asymptotic approximations to the finite-sample distribution of the estimator, along the lines of Rothenberg (1984). Another approach, proposed by Staiger and Stock (1997), is to derive an alternative first-order asymptotic theory by linking the key parameter $\Pi$ to the sample size.

Both of these approaches can be motivated by re-writing the IV estimator $\widehat{\theta}$ as a function of some pivotal statistics (that is, statistics whose distribution

[2]For more general treatments, see Phillips (1989) or Staiger and Stock (1997).

is free from any parameters). This is straightforward in the univariate case, $p = 1$. Define the two independent standard normal variates:

$$z = \begin{pmatrix} z_e \\ z_v \end{pmatrix} = \begin{pmatrix} (Z'Z)^{-1/2} Z'e/\sigma_e \\ (Z'Z)^{-1/2} Z'v/\sigma_v \end{pmatrix} \sim N(0, I_{2k}),$$

and their linear combination $z_u = (Z'Z)^{-1/2} Z'u/\sigma_u = z_e\sqrt{1-\rho^2} + z_v\rho$. Also, let

$$\mu_T = (Z'Z)^{1/2}\Pi/\sigma_v. \tag{8}$$

This quantity is known as the concentration parameter (Anderson 1977). Then, dropping the subscript of $\mu_T$ for simplicity, the IV estimator (6) in the one-parameter case can be written as:

$$\widehat{\theta}_{IV} - \theta_0 = \frac{(z_v+\mu)'\, z_u\, \sigma_u/\sigma_v}{(z_v+\mu)'\,(z_v+\mu)} = \frac{(z_v+\mu)'\,(z_e\,\frac{\sigma_{u,v}}{\sigma_v} + z_v\,\lambda)}{(z_v+\mu)'\,(z_v+\mu)}. \tag{9}$$

The above expression highlights the dependence of the finite sample distribution of the IV estimator on nuisance parameters, as well as its departure from normality. Since the random vectors $z_v$ and $z_e$ are independent, the distribution is a location-scale mixture of normals, and in special cases it can be represented as a doubly non-central $t$-distribution, see Phillips (1984). Evidently, the variability of $\widehat{\theta}_{IV}$ and its departure from normality depend on the modulus of $\mu$. When we let $||\mu|| \to \infty$, the normalized IV estimator becomes:

$$\sqrt{\mu'\mu}\,\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0,\ \frac{\sigma_u^2}{\sigma_v^2}\right)$$

Expression (9) also allows us to make statements about the finite-sample bias of the IV estimator of $\theta$. This clearly depends on $\lambda$, $\mu$ and the number of instruments $k$, but when $\mu'\mu/k$ is large, it is (approximately) inversely related to $\mu'\mu/k$:

$$B_{IV} = \mathrm{E}(\widehat{\theta} - \theta_0) = \mathrm{E}\left[\frac{(z_v+\mu)'z_v\lambda}{(z_v+\mu)'(z_v+\mu)}\right] \approx \frac{\mathrm{E}\left[(z_v+\mu)'z_v\lambda\right]}{\mathrm{E}\left[(z_v+\mu)'(z_v+\mu)\right]} = \frac{\lambda}{1+\mu'\mu/k}.$$

Similar calculations would show the approximate OLS bias to be $B_{OLS} \approx \frac{\lambda}{1+\mu'\mu/T}$, which is intuitive, since OLS can be thought of as IV with as many instruments as there are observations.

Finally, since the standard error of $\widehat{\theta}$ is the most commonly used measure of its precision and similarly the t-statistic is the most popular method of inference in the regression context, it is useful to look more closely into their finite sample properties. Though it is straightforward to derive these expressions directly, application of Staiger and Stock (1997, Theorem 1) yields:

$$\widehat{\sigma}_u^2 = \sigma_u^2 + \left(\widehat{\theta}_{IV} - \theta_0\right)\left[\left(\widehat{\theta}_{IV} - \theta_0\right) - 2\lambda\right]\sigma_v^2 \tag{10}$$

$$s.e.(\hat{\theta}) = \widehat{\sigma}_u(\widehat{\Pi}'Z'Z\widehat{\Pi})^{-1/2} = \frac{\widehat{\sigma}_u/\sigma_v}{[(z_v+\mu)'(z_v+\mu)]^{1/2}} \tag{11}$$

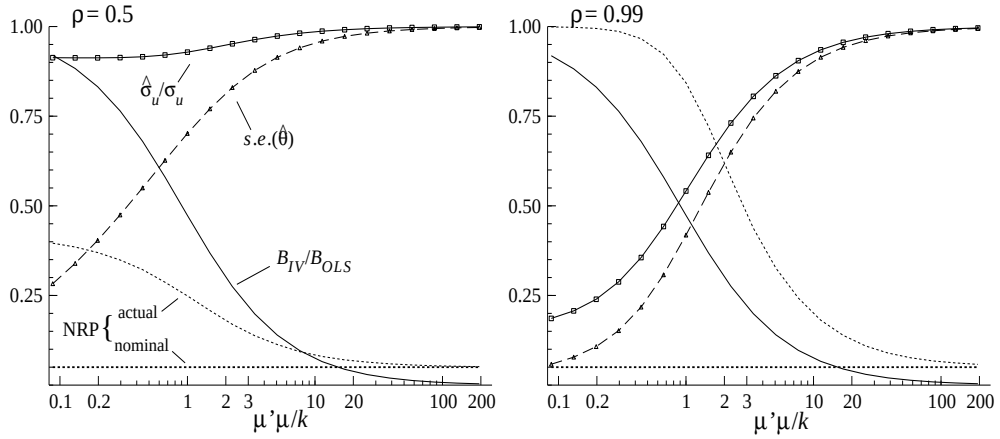


Figure 1: Mean biases of IV estimators of $\theta$ (relative to OLS bias), $\sigma_u$ and $s.e.(\widehat{\theta})$ (relative to their true values), and the actual Null Rejection Probability (NRP) of a 5% level t-test on $\theta_0$. The number of instruments is $k = 10$. $\mu'\mu$ is the concentration parameter (see eq. (12)), which is a scalar here and thus coincides with its minimum eigenvalue, $\mu^2_{min}$. (Logarithmic scale on the x-axis).

The t-statistic is simply $t = \frac{\widehat{\theta}_{IV}-\theta_0}{s.e.(\widehat{\theta})}$. We see from (10) that the structural variance is under-estimated whenever the IV estimator lies between 0 and twice the OLS bias. Exact calculation of $\mathrm{E}(\widehat{\sigma}_u)$ using (10) reveals that this happens more often than not, namely $\widehat{\sigma}_u$ is biased downwards in finite samples. From (11) we expect the estimate of the standard error also to be below its true asymptotic value of $\sigma_u/\sqrt{\Pi'\Sigma_{ZZ}\Pi}$. Finally, with regards to the t-statistic, we expect it to dominate its assumed t-distribution, due to a positive non-centrality in the numerator (arising from the finite sample bias of $\widehat{\theta}_{IV}$) and an under-estimated denominator. Hence, we expect the t-test to over-reject the null hypothesis $\mathrm{H}_0 : \theta = \theta_0$.

In Figure 1, we plot $B_{IV}/B_{OLS}$, $\widehat{\sigma}_u/\sigma_u$, $s.e.(\widehat{\theta})/(\sigma_u/\sqrt{\Pi'\Sigma_{ZZ}\Pi})$ and the actual null rejection probability (NRP) of a nominal 5% level t-test on $\theta_0$, against $\mu'\mu/k$. We do this for two benchmark levels of endogeneity, $\rho = 0.5$ and $\rho = 0.99$, following the convention in the literature. Our intuition is corroborated by the graphs. The relative bias is falling in $\mu'\mu/k$, and is unaffected by the degree of endogeneity. For small values of $\mu'\mu/k$ we observe large biases in the variance estimators too, and considerable over-rejection of the t-test. These effects are more pronounced the higher the endogeneity $\rho$.

Before concluding this section, we point out the most remarkable feature of the above results: namely, the sample size $T$ does not enter explicitly in the distribution of any of the above statistics, except through the concentration parameter $\mu_T$ (we dropped its dependence on $T$ earlier, for convenience). That is, 'small sample' problems arise when $||\mu_T||$ is small, and not necessarily when $T$ itself is small. Of course, the above analysis was highly stylized, based on unrealistic assumptions, such as the strong exogeneity of the instruments,

and the conditional normality of the endogenous variables, none of which is satisfied in practice. This analysis is usually justified as an approximation, in which case an explicit dependence on the sample size arises directly (e.g., by approximating covariance matrices by their empirical counterparts). But the important message is that when the sample size is large enough for the intuition of this analysis to be relevant, it is the concentration parameter that determines how informative the data is for our parameters of interest.

**More regressors** The presence of exogenous regressors $X$, say, in the structural equation (4) doesn't pose any additional challenge. The above analysis holds exactly if we replace $W = (y, Y, Z)$ by the residuals of their projection onto $X$, namely, $W^{\perp} = (I - X(X'X)^{-1}X')W$. However, the exogenous coefficient estimators will be affected by partial or weak identification of the endogenous parameters, and can even be inconsistent when $X$ correlates with $Y$ (except through $Z$), see Choi and Phillips (1992) and Staiger and Stock (1997).

When the number of endogenous regressors is $p > 1$, the concentration parameter (8) is a matrix of dimension $p$:

$$\mu'\mu = T\,\Sigma_{vv}^{-1/2}\Pi'\,\Sigma_{ZZ}\,\Pi\Sigma_{vv}^{-1/2}. \tag{12}$$

Thus, partial identification arises whenever $\text{rank}(\mu) = \text{rank}(\Pi) < p$, or equivalently, when some of its eigenvalues are zero (Phillips 1989). Moreover, the eigenvectors corresponding to the non-zero eigenvalues give the linear combinations of the structural parameters that are identified, see examples below.

In contrast, generic identification corresponds to the situation $\text{rank}(\mu) = p$. In this case, the minimum eigenvalue of the concentration parameter, denoted $\mu^2_{min}$ or simply $\mu^2$ when $p = 1$, could serve as an index of the strength of identification (Stock and Yogo 2003). In particular, empirical identification arises when $\mu^2_{min}$ is large, e.g., in the sense of Stock and Yogo (2003), while a small $\mu^2_{min}$ implies weak identification. Again, a singular value decomposition of $\mu$ will reveal the parameter combinations that are well-identified and those that are weakly identified.

## Checking identification

The above analysis demonstrated why the usual rank condition for identification is not sufficient for reliable estimation and inference in finite samples. This emphasizes the need to check identification prior to GMM estimation.

Several procedures are available for testing the null hypothesis of underidentification against the alternative of generic identification. These procedures amount to testing for rank deficiency in the coefficient matrix $\Pi$ in the first-stage regression (5), or in the matrix $\mu$ directly. In models, with one endogenous variable ($p = 1$), this can be done with a simple F-test of joint significance in the first-stage regression of the instruments that are additional to

any exogenous regressors. When $p > 1$, the hypothesis of under-identification can be tested using the reduced rank regression technique, developed by Anderson and Rubin (1950).

In forward-looking models where autocorrelated errors typically arise by construction, and heteroscedasticity of the residuals cannot be ruled out a priori, the standard likelihood ratio 'trace statistic' for reduced rank cannot be used, and one must resort to more robust methods, e.g., Cragg and Donald (1997), Robin and Smith (2000) or Kleibergen and Paap (2003).[3]

However, it must be emphasized that none of these tests can distinguish between weak and strong identification, but they have power even in situations where empirical identification is still weak (Staiger and Stock 1997). There are two alternative diagnostic tools that offer a solution to this problem. One is the Hahn and Hausman (2002) test, which is a Hausman-type test of the null hypothesis of strong identification. The other is the approach of Stock and Yogo (2003), which is based on the minimum eigenvalue of the concentration parameter $\mu^2_{min}$. Given a practical criterion for judging empirical identification, such as maximum tolerable bias of an estimator, or size of a given test, one may derive a critical region for $\mu^2_{min}$, in which identification will be deemed weak. Then, weak identification can be checked by testing that $\mu^2_{min}$ is in that region (rather than exactly 0).[4] Unfortunately, those tests only apply to the case of the linear IV model with homoscedastic and serially uncorrelated residuals. Hence, they are not applicable to forward-looking RE models.[5]

A related approach is to compute an empirical estimate of the concentration parameter, $\widehat{\mu}$, say, by replacing population moments in (12) by their empirical counterparts. Of course, it should be pointed out that the concentration parameter is not consistently estimable, in the sense that there is no estimator with the property $\widehat{\mu} - \mu \xrightarrow{p} 0$. Instead, $\widehat{\mu} - \mu = O_p(1)$, but can be asymptotically unbiased, meaning that as the sample size grows, $\widehat{\mu} - \mu$ approaches a mean-zero non-degenerate distribution. For instance, under the assumptions of section II, $\widehat{\mu} - \mu = z_v + o_p(1) \overset{a}{\sim} N(0, I_{kp})$.

In the context of forward-looking models, an important drawback of the above identifiability pre-tests that focus on the correlation between endogenous regressors and instruments is that they cannot distinguish between identification and mis-specification of a model. Mavroeidis (2003) shows how identification of a forward-looking model can be achieved through dynamic mis-specification. In order to separate the identification analysis from mis-

[3]In the notation of section II, the limiting variance of $T^{-1}Z'[u,\ v]$, $V(\theta)$ say, no longer has the convenient Kroneker form $\Sigma_{UU} \otimes \Sigma_{ZZ}$. Under standard regularity conditions, $Avar(T^{-1}Z'v) = V_{22}$ is consistently estimable by a HAC estimator, and that can be used to form an identification test, see Mavroeidis (2002, section 2.3) for more discussion.

[4]The above-mentioned test statistics for reduced rank can be used for this more general hypothesis, at the expense of yielding conservative inference (i.e., they are biased towards diagnosing weak identification too often). Nevertheless, Stock and Yogo (2003) argue that their approach is informative.

[5]Extension of the Stock and Yogo (2003) approach to the linear GMM framework is possible, but still in progress.

specification, one needs to utilize the dynamic structure and the rational expectations condition of the model, in the style of Pesaran (1987).

However, the identification analysis of Pesaran (1987) is confined to the conditions for generic identification. This helps uncover pathological situations in which the rank condition fails, but is not sufficient to discuss empirical identification. This can be done by deriving a measure of identification which is conditional on the correct specification of the model, e.g., replacing $\Sigma_{ZZ}, \Sigma_{vv}$ and $\Pi$ in (12) by estimators that are restricted by the time-series structure of the driving variables and the requirement that the forward-looking model (1) be correctly specified. When the resulting measure of the minimum eigenvalue of the concentration parameter, $\widehat{\mu}^2_{\min}$, is very small (e.g. less than 1) while the number of instruments is large, we may conclude that the model is weakly identified (with the caveat of estimation uncertainty). If this contradicts the conclusions drawn from statistical pre-tests of identifiability, then it has potential implications for the specification of the model.

# III Forward-looking models

When estimated by GMM, forward-looking rational expectations models of the form (1) are a special case of the generic structural model (4), where the endogenous regressors $Y$ include leads of the endogenous variables, and the instrument set $Z$ contains lags of the endogenous and the driving variables $(e_t)$.

However, there are some important differences with the stylized example in the previous section, which need to be pointed out. First, forward-looking models are dynamic, and impose more structure on the reduced form parameters, $\Pi$, which are typically linked to the parameters of interest $\theta$. Second, the instruments are not strongly exogenous. And third, both the structural as well as the reduced-form errors $u$ and $v$, are autocorrelated by construction.

Despite these differences, the intuition of the static analysis of the previous section offers considerable insights into the finite sample properties of GMM estimators in forward-looking models.[6] In those models too, the rank condition for identification is not sufficient for reliable estimation and inference. Therefore, the identification analysis proposed in Pesaran (1987, chapter 6) can only serve as a starting point.

In this section, we will offer some illustration of the implications of partial and weak identification on the estimates of a forward-looking parameter in an equation of the form (1).

[6]See Mavroeidis (2002) for some relevant asymptotic theory and extensive Monte Carlo evidence.

## Identification analysis

The identification analysis of a forward-looking rational expectation model (1) requires knowledge of the second moments of the data. These cannot be derived from the structural equation (1) alone, since it is an incomplete description of the local DGP. Instead, we need to provide a completing model for the forcing variables $e_t$, and then solve the system using rational expectations. Pesaran (1987) showed that the factors governing the rank condition of identification are: (a) the specification of the information set $\mathcal{F}_t$; (b) the type of solution to the rational expectations model ('forward' or 'backward'); and (c) the dynamics of the driving process $e_t$.

Concerning the information set, we follow the convention in the literature and assume that it includes *at least* all contemporaneous and past information on the endogenous and forcing variables (Binder and Pesaran 1995).

The conditions for existence and uniqueness of non-explosive solutions to rational expectations models were provided by Blanchard and Kahn (1980). In short, when (without loss of generality) the forcing variables $e_t$ are not Granger-caused by the endogenous variables $y_t$, a non-explosive solution to model (1) exists when the number of explosive roots in the lag polynomial $I-\beta L^{-1}-\gamma L$, does not exceed the number of endogenous variables. In a single-equation partial adjustment model, this amounts to at most one explosive root. When there is exactly one explosive root, the solution is unique, and this is sometimes referred to as the 'forward' solution. When there are no explosive roots, the 'backward' solution is non-unique and takes the generic form:

$$y_t = \frac{1}{\beta}y_{t-1} - \frac{\gamma}{\beta}y_{t-2} - \frac{1}{\beta}e_{t-1} + \xi_t \tag{13}$$

where $\xi_t$ is an arbitrary Martingale Difference Sequence with respect to $\mathcal{F}_{t-1}$ satisfying $\mathrm{E}(\xi_t|\mathcal{F}_{t-1}) = 0$.[7] It can be shown that the unique forward solution is always nested within the class of backward solutions, i.e. it can be derived by parametric restrictions on (13).[8]

The conditions for the (generic) identification of the parameters of the structural equation (1) will depend on the type of solution, so we analyze each case separately.

**Backward solutions** Provided there are no common factors in the lag structure of $y_t$ and that of $(e_{t-1}+\xi_t)$ in the solution equation (13) that would reduce it to the forward solution (see below), the rank condition for identification of $\beta$ (and $\gamma$) will always be satisfied, irrespective of the dynamics of $e_t$. This is most easily seen in the pure forward-looking model (1), with $\gamma = 0$. As an

[7] This shock may correlate with the innovation in the forcing variable, e.g., $\xi_t = a\,(e_t - \mathrm{E}(e_t|\mathcal{F}_{t-1})) + \varsigma_t$. The orthogonal part $\varsigma_t$ is referred to as a 'sunspot shock'. This shock cannot, in general, be identified without strong assumptions, such as the requirement that the rational expectations model be 'exact', (Hansen and Sargent 1991).

[8] Pesaran (1987, pp. 143-144).

example, consider the New Phillips curve model, with $e_t = \lambda\, s_t$. The GMM estimating equation is:

$$\pi_t = \beta\, \pi_{t+1} + \lambda\, s_t + u_t$$

where $u_t = -\beta\, [\pi_{t+1} - \mathrm{E}(\pi_{t+1}|\mathcal{F}_t)]$, and, when $|\beta| > 1$ the reduced-form solution (13) is:

$$\pi_t = \frac{1}{\beta}\, \pi_{t-1} - \frac{\lambda}{\beta}\, s_{t-1} + \xi_t.$$

Using the instruments $(s_t, s_{t-1}, \pi_{t-1})$, the first-stage regression for the endogenous regressor $\pi_{t+1}$ is:

$$\pi_{t+1} = \frac{1}{\beta^2}\pi_{t-1} - \frac{\lambda}{\beta}s_t - \frac{\lambda}{\beta^2}s_{t-1} + v_t$$

with $v_t = \xi_{t+1} + \beta^{-1}\, \xi_t$. Since $s_t$ is effectively an 'exogenous' regressor, the identification analysis can be carried out by orthogonalizing the remaining variables to $s_t$, namely:[9]

$$\pi^{\perp}_{t+1} = \frac{1}{\beta^2}\pi^{\perp}_{t-1} - \frac{\lambda}{\beta^2}s^{\perp}_{t-1} + v_t$$

where $w^{\perp}_t$ means $w_t - \mathrm{E}[w_t|\, s_t]$. In the standard IV notation of section II, $\Pi = \frac{1}{\beta^2}(1,\ -\lambda) \neq 0$, so that we can conclude that the parameters $(\beta, \lambda)$ are generically identified *irrespective* of the process generating $s_t$.

However, as we argued above, the question of empirical identification can be answered by looking at the concentration parameter. In this case, this is the ratio of the predictable $(\frac{1}{\beta^2}\pi^{\perp}_{t-1} - \frac{\lambda}{\beta^2}s^{\perp}_{t-1})$ relative to the unpredictable $(v_t)$ variation in $\pi^{\perp}_{t+1}$. This will, in turn, depend on the variance of the sunspot shock, albeit in a rather complicated manner. The variance of $\xi$ is positively related to the noise $(\mathrm{var}(v_t))$, but also to the signal, since it increases the variance of the instrument $\pi^{\perp}_{t-1}$. It seems impossible to be more precise without specifying a completing model for $s_t$.[10] Thus, we see that even in the case where generic identification is guaranteed, the specification of the completing process will generally be informative about the extent of empirical identification.

Another reason for concern is the qualification we made earlier, namely that there should be no common factors in the lag polynomials in the solution equation (13) that would reduce it to the forward solution. This is the case we turn to next.

[9] In writing $v^{\perp}_t = v_t$ we have implicitly assumed that $\xi_t$ is a pure sunspot shock, and hence orthogonal to $(s_t, s_{t-1}, \ldots)$. If we relax that assumption, we introduce the possibility of an even higher degree of over-identification (more lags of $s_t, \pi_t$ being relevant instruments), see Mavroeidis (2002, section 4.2.1). We do not discuss this for clarity, and also since the emphasis is on pathological cases of weak identification.

[10] For instance, when $s_t \sim iid(0, \sigma^2_s)$ and orthogonal to $\xi_t$, then we can derive $\mu^2_{min} = \frac{1}{\beta^2(\beta^2-1)}\left(1 + \lambda^2\frac{\sigma^2_s}{\sigma^2_\xi}\right)$, showing that the strength of identification is falling with the variability of the sunspot shock. Also, identification is stronger the closer $\beta$ is to 1.

**Forward solution** The rational expectations model (1) has a unique solution when the polynomial $\beta z - \gamma z^{-1} = 1$ has exactly one explosive root. Let the stationary root be $\delta = \frac{1-\sqrt{1-4\beta\gamma}}{2\beta} < 1$ so that the explosive root is $\gamma/(\beta\delta) > 1$. Then, the unique solution to (1) is given by (see, for instance, Pesaran 1987, pp.108–109)

$$y_t = \delta\, y_{t-1} + \frac{\delta}{\gamma} \sum_{j=0}^{\infty} \left(\frac{\delta\beta}{\gamma}\right)^j \mathrm{E}(e_{t+j}|\mathcal{F}_t). \tag{14}$$

In the pure forward-looking case, $\gamma = 0$ implies $\delta = 0$ and $\delta/\gamma = 1$ (by l'Hopital).

Suppose the forcing variable is of the form $e_t = \lambda(L)x_t + \epsilon_t$, where $\lambda(L)$ is a lag polynomial of order $p$, and $x_t$ is a covariance stationary process, not Granger-caused by $y$, that admits an AR($q$) representation, and $\epsilon_t$ is a mean innovation process w.r.t. $x_t, y_t$.[11] Then, a sufficient condition for the identification of the forward-looking parameter $\beta$ (and also $\gamma$ if present, as well as the $\lambda$'s), using GMM with at least $p+2$ lags of $x_t$ as instruments, is $q > p+1$ (Pesaran 1987, Proposition 6.2). In other words, the forcing variables must have more dynamics than what is already included in the structural model.

Another way of putting the above result is that the unique solution to the structural model, which would be of the form $y_t = \delta\, y_{t-1} + \alpha(L)\, x_t$, must not be nested within that model, $y_t = \beta\, y_{t+1} + \gamma\, y_{t-1} + \lambda(L)\, x_t$. If the polynomials $\alpha(L)$ and $\lambda(L)$ were of the same order, and their coefficients were unrestricted, then there would be more structural parameters $(\beta, \gamma, \lambda_i)$ than estimable reduced-form parameters $(\delta, \alpha_i)$, so the former would be un-identified (on the order condition).

As an example, consider a pure forward-looking version of the New Phillips Curve (3) with $\gamma = 0$ and $e_t = \lambda\, s_t + \epsilon_t$, where $\epsilon_t$ is an exogenous inflation innovation. The theoretical framework of Galí and Gertler (1999) provides an economic interpretation to the parameters $(\beta, \lambda)$. The former is a discount factor and therefore it is restricted to lie between 0 and 1. When $\beta$ is strictly less than 1, the model has a unique forward solution, and hence the rank condition for identification is satisfied if $s_t$ has at least second-order dynamics. Of course, empirical identification depends on the nature of the dynamics of $s_t$, as well as its relation to the endogenous variable $\pi_t$.

Consider, for instance, the complete rational expectations model:

$$\pi_t = \beta \mathrm{E}(\pi_{t+1}|\mathcal{F}_t) + \lambda s_t + \epsilon_t \tag{15}$$

$$s_t = \rho_1 s_{t-1} + \rho_2 s_{t-2} + \zeta_t \tag{16}$$

Uniqueness of the solution ($|\beta| < 1$) would imply that the first-stage regression

[11] This represents information known to the agents when forming their decisions, but not to the econometrician, i.e. a measure of the incompleteness of the structural equation. Such a process is always empirically plausible. Otherwise, the absence of $\epsilon_t$ together with the forward solution (14) would imply that the joint distribution of $y_t$ and $x_t$ is stochastically singular.

for $\pi_{t+1}$ would be of the form:

$$\pi_{t+1} = \alpha_1 \, s_t + \alpha_2 \, s_{t-1} + v_t, \tag{17}$$

with $v_t = \epsilon_{t+1} + \alpha_1 \, \zeta_{t+1}$, and $(\alpha_1, \alpha_2)$ are given in the appendix. So, the only relevant instrument in this case (beyond $s_t$ which is included as an exogenous regressor) is $s_{t-1}$, or better, the residual of its projection onto $s_t$, $s^{\perp}_{t-1}$. Therefore, the concentration parameter is (see appendix)

$$\mu^2 = \frac{\alpha_2^2 \, \mathrm{var}(s^{\perp}_{t-1})}{\mathrm{var}(v_t)} = \frac{\alpha_2^2 \, \sigma_\zeta^2}{(1-\rho_2^2) \, \left(\alpha_1^2 \, \sigma_\zeta^2 + 2 \, \alpha_1 \, \sigma_{\epsilon\zeta} + \sigma_\epsilon^2\right)}. \tag{18}$$

This expression reveals clearly that a 'statistically significant' second order dynamic adjustment in $s_t$ is by no means sufficient to guarantee empirical identification. It is true that the strength of identification is increasing in $|\rho_2|$, other things equal.[12] It is unambiguously increasing in $\sigma_\zeta^2$, too, since the latter contributes more to the signal than to the noise. But, importantly, identification is decreasing in the exogenous variability in inflation. This is particularly relevant for the identifiability of monetary models like (2) and (3), as we argue below.

Based on the original data and results reported by Galí and Gertler (1999), the estimated value of (18) is of the order $10^{-4}$, lending support to the view that the New Keynesian Phillips curve of Galí and Gertler (1999) is weakly identified on their information set, see Mavroeidis (2003). However, this conclusion is conditional on the model (15) being correctly specified and having a unique forward solution, as their reported parameter estimates suggest. Otherwise, identification may arise through omitted dynamics in (15), see Mavroeidis (2003) for details.

In sum, the main sources of weak identification in forward-looking models like (1) are that: (i) the forcing variables have limited dynamics, and/or (ii) the un-predictable variation in future endogenous variables is large relative to what is predictable on the available instruments. Additionally, when the model admits a backward solution, while its forward solution is such that it would be under-identified, weak identification can result when the lag polynomials in the solution of the model are close to having a common factor.

## Forward-looking Taylor rules

Empirical estimates of forward-looking Taylor rules of the form (2) also allow for additional dynamics in the interest rate, because lags of the interest rate appear to be statistically significant. This generalization is referred to as

[12] In the appendix, we show that $\alpha_1$ and $\alpha_2$ depend on $\rho_1$ and $\rho_2$ as well as the structural parameters $(\beta, \lambda)$. So, identification will also depend on the true values of the structural parameters. If instead of keeping $(\alpha_1, \alpha_2)$ fixed at their true values, we choose to talk about identification of a particular $(\beta_0, \lambda_0)$, then the comparative statics in equation (18) will be slightly more involved.

'interest rate smoothing'. Here, we focus on the econometric specification of Clarida, Galí, and Gertler (2000):

$$r_t = \rho\, r_{t-1} + (1-\rho)\, [\beta\, \mathrm{E}(\pi_{t+1}|\mathcal{F}_t) + \gamma\, \mathrm{E}(x_{t+1}|\mathcal{F}_t)] + \epsilon_t \tag{19}$$

where $r_t$ and $\pi_t$ are understood to be in deviations from their neutral level and target values respectively. Gathering the target variables in a vector $w_t = (\pi_t, x_t)$, and letting $\theta = (1-\rho)(\beta, \gamma)'$, we may re-write (19) as:

$$r_t = \rho\, r_{t-1} + \theta' \mathrm{E}(w_{t+1}|\mathcal{F}_t) + \epsilon_t \tag{20}$$

Equation (19) is cast into a GMM regression

$$r_t = \rho\, r_{t-1} + \theta'\, w_{t+1} + u_t \tag{21}$$

with $u_t = \epsilon_t - \theta'\, (w_{t+1} - \mathrm{E}(w_{t+1}|\mathcal{F}_t))$. Then, $(\beta, \gamma, \rho)$ can be estimated using instruments in the $t$-dated information set, which should, in principle, include contemporaneous values of $\pi_t$ and $x_t$. In practice, researchers only use lagged information, presumably due to measurement lags. In our identification analysis, we will conform with this common practice.

To discuss identification, we look at the concentration parameter, as before. This, in turn, requires knowledge of the reduced form of the model. To derive this, we need to provide a completing model for inflation and the output gap (the monetary transmission mechanism).

There are two ways we can proceed. One is to use a backward-looking completing model of the transmission mechanism, such as a Vector Autoregressive Distributed lag model (VAD) for $w_t$ given the interest rate $r_t$ and any additional variables that might be used as instruments.This system will only serve to derive the first-stage regression, and therefore it need not have any structural interpretation. One objection to this approach is that a constant-parameter VAD is an unrealistic econometric model for inflation and the output gap over any long period of time, because it does not address the Lucas (1976) critique.

An alternative approach would be to embed equation (19) in a complete forward-looking model for the three variables of interest $(\pi_t, x_t, r_t)$. The solution of that system would be the restricted reduced form that would be used to conduct the identification analysis. Not only does this approach address the Lucas critique, but it also makes the analysis of identification of equation (19) consistent with the theoretical framework that is used to justify it, see Clarida, Galí, and Gertler (2000, Section 4).

However, we notice that, even when we postulate a forward-looking model for inflation and the gap, provided this model is a linear multivariate rational expectations model, any solution to the system will be a particular restricted reduced-form model for $(w_t', r_t)'$. Hence, for the purposes of empirical identification analysis, it suffices to look at the unrestricted reduced form for $w_t$ as a completing model.

**Identification analysis using a backward-looking completing model**

We assume that the Taylor rule is an accurate description of interest rate dynamics, and that the unrestricted reduced form of the endogenous variables can be represented by a Vector Autoregressive Distributed lag model of orders $n, m$, say:

$$w_t = \sum_{i=1}^{n} A_i \, w_{t-i} + \sum_{j=1}^{m} b_j \, r_{t-j} + \eta_t. \tag{22}$$

$A_i$ are $2 \times 2$ matrices, $b_j$ are $2 \times 1$ vectors, and $\eta_t$ are iid innovations that can be assumed to be uncorrelated to the (exogenous) policy shock $\epsilon_t$.

To compute the concentration parameter (12), we shall derive the first-stage regression, in the prototype form (5). We also need to account for the exogenous regressor $r_{t-1}$ in the structural equation, by orthogonalizing the remaining instruments to it, as discussed on p. 9. From equation (22) and (20) we can derive the forecasting regression of $w_{t+1}$ on the information set $\mathcal{F}_{t-1}$ (see appendix)

$$\begin{aligned} w_{t+1} &= (I + \delta b_1 \theta') \left( \sum_{i=1}^{n} (A_1 A_i + A_{i+1}) w_{t-i} + \sum_{j=2}^{m} (A_1 b_j + b_{j+1}) r_{t-j} \right) \\ &\quad + (I + \delta b_1 \theta')(A_1 b_1 + b_2 + b_1 \rho) r_{t-1} \\ &\quad + \eta_{t+1} + (I + \delta b_1 \theta') A_1 \, \eta_t + \delta b_1 \epsilon_t \end{aligned} \tag{23}$$

where $\delta = 1/(1 - \theta' b_1)$, and we have assumed that $\theta' b_1 \neq 1$.[13]

Next, define the vector of all relevant instruments $Z_t = (w'_{t-1}, \ldots, w'_{t-n}, r_{t-2}, \ldots, r_{t-m})'$. The first stage regression residual can be seen from equation (23) to be $v_t = \eta_{t+1} + (I + \delta b_1 \theta') A_1 \, \eta_t + \delta b_1 \epsilon_t$. Thus, the first-stage regression coefficient, $\Pi$, and the variance of $v_t$ are given by

$$\begin{aligned} \Pi' &= (I + \delta b_1 \theta')(A_1^2 + A_2, \ldots, A_1 A_{n-1} + A_n, A_1 A_n, \\ &\quad A_1 b_2 + b_3, \ldots, A_1 b_{m-1} + b_m, A_1 b_m) \end{aligned} \tag{24}$$

$$\Sigma_{vv} = [I + (I + \delta b_1 \theta') A_1] \Sigma_{\eta\eta} [I + (I + \delta b_1 \theta') A_1]' + \delta^2 b_1 b_1' \sigma_\epsilon^2. \tag{25}$$

The final ingredient to compute the concentration parameter is the variance of the instruments, $Z_t$, corrected for the exogenous regressor $X_t = r_{t-1}$. Using the notation $Z_t^\perp = Z_t - \mathrm{E}(Z_t | X_t)$ as before, define $\Sigma_{ZZ}^\perp = \mathrm{var}(Z_t^\perp) = \mathrm{var}(Z_t | X_t) = \Sigma_{ZZ} - \Sigma_{ZX} \Sigma_{XX}^{-1} \Sigma_{ZX}'$. Hence, $\Sigma_{ZZ}^\perp$ is simply a function of the unconditional second moments of the data. Those can be readily derived from the reduced form of the entire system, which is a VAR of order $\max(n, m)$, as shown in the appendix.

[13] If $\theta' b_1 = 1$, no solution to the policy rule (20) exists under Rational Expectations. This degenerate case is rather implausible, since typically $\theta > 0$ and $b_1 < 0$, as real interest rates should correlate negatively with future inflation and output in well-specified models of the transmission mechanism.

Unlike the example of the New Keynesian Phillips curve in the previous section, an analytical treatment of the concentration matrix and its minimum eigenvalue $\mu^2_{min}$ is intractable. Instead, we can study the benchmark cases of under-identification, in order to characterize the pathological subset of the parameter space where identification is lost.

Let $\Omega$ denote the parameter space, containing all the possible values the parameters $\beta, \gamma, \rho, \sigma_\epsilon, \{A_i\}_{i=1}^n, \{b_j\}_{j=1}^m$ and $\Sigma_{\eta\eta}$ can take. The non-identification region $\Omega_0$ is a subset of $\Omega$ that contains all the values of the parameters for which $\mu^2_{min} = 0$.

There are two potential sources of under-identification. The classic one is when $\Pi$ is of reduced rank. Another possibility, which is often overlooked or assumed away, is that the exogenous regressor $r_{t-1}$ is perfectly collinear with the optimal instruments $\Pi' Z_t$. We will argue that such a degenerate case is, in fact, not implausible in the context of monetary policy.

Consider first the rank of $\Pi$, given by (25). Note that the condition $\theta' b_1 \neq 1$ implies that the matrix $(I + \delta b_1 \theta')$ is non-singular (see appendix). Hence, the rank of $\Pi$ depends only on the parameters $\{A_i\}_{i=1}^n, \{b_j\}_{j=2}^m$. In particular, a necessary condition for generic identification is

$$\text{rank}(A_1^2 + A_2, \ldots, A_1 A_{n-1} + A_n, A_1 A_n, A_1 b_2 + b_3, \ldots, A_1 b_{m-1} + b_m, A_1 b_m) = 2. \tag{26}$$

In general, under-identification occurs if there exists a linear combination of the two endogenous regressors $w_{t+1}$ that is not predictable on $\mathcal{F}_{t-1}$ beyond $r_{t-1}$, e.g., $d \in \Re^2$ such that $d' w^\perp_{t+1} = d' v_t$. This happens if all the matrices $\{A_1 A_i + A_{i+1}\}_{i=1}^{n-1}$, $A_1 A_n$, $\{A_1 b_j + b_{j+1}\}_{j=2}^{m-1}$ and $A_1 b_m$ have common and non-empty kernels. All the values of $A_i$ and $b_j$ that satisfy this condition lie in the non-identification region.

A particular example is when there exists a linear combination of inflation and the gap that has no dynamics beyond the first and second lag of $r_t$, i.e. $d' A_i = 0$ for all $i$ and $d' b_j = 0$ for $j > 2$ in (22). It is straightforward to verify that the necessary condition (26) fails in this case. However, such degeneracy is not necessary for (26) to fail. This can happen even when $A_1$ is non-singular, for instance.

Partial identification can also occur if the exogenous regressor $r_{t-1}$ is perfectly collinear with the optimal instruments.The restrictions on the parameters that would induce such collinearity can be derived from (23) and (20). Lag (20) one period, substitute for $\text{E}(w_t|\mathcal{F}_{t-1})$ from (22) and re-arrange to get $r_{t-1}$ as a function of $Z_t$. Noting that $(I + \delta b_1 \theta')$ is non-singular, the optimal instruments can be derived from (23) as $Z_t^{opt} = (I + \delta b_1 \theta')^{-1} \Pi' Z_t$. The two equations are

$$r_{t-1} = \sum_{i=1}^{n} \delta \theta' A_i \, w_{t-i} + \delta(\rho + \theta' b_2) \, r_{t-2} + \sum_{j=3}^{m} \delta \theta' b_j \, r_{t-j} + \epsilon_{t-1} \tag{27}$$

$$Z_t^{opt} = \sum_{i=1}^{n} (A_1 A_i + A_{i+1}) w_{t-i} + \sum_{j=2}^{m} (A_1 b_j + b_{j+1}) r_{t-j} \tag{28}$$

Perfect collinearity means there is a linear combination of $Z_t^{opt}$ that is identically equal to $r_{t-1}$ for all $t$, or, alternatively, that the first canonical correlation between $Z_t^{opt}$ and $r_{t-1}$ is unity. Let the linear combination $d'Z_t^{opt}$ denote the first canonical variate of $Z_t^{opt}$ with $r_{t-1}$. To derive the necessary restrictions on the parameters for perfectly collinearity, premultiply (28) by $d'$ and equate the resulting right-hand side coefficients with those of (27). Upon re-arrangement, the restrictions can be written recursively as follows:

$$\begin{aligned} d'A_{i+1} &= (\delta\theta' - d'A_1)A_i, \quad i = 1, \ldots, n-1, \\ (\delta\theta' - d'A_1)A_n &= 0, \\ d'b_3 &= (\delta\theta' - d'A_1)b_2 + \delta\rho, \\ d'b_{j+1} &= (\delta\theta' - d'A_1)b_j, \quad j = 3, \ldots, m-1, \\ (\delta\theta' - d'A_1)b_m &= 0. \end{aligned} \tag{29}$$

Note also the necessity of $\epsilon_t = 0$ for all $t$, that is, the absence of a 'monetary policy shock'. This is interesting because it suggests that, in certain cases, the presence of such a shock will help identify an otherwise un-identified model. Suppose, for instance, the transmission mechanism is such that a policy rule like (21) would be optimal (in the sense of minimizing a particular loss function that penalizes deviations of inflation and output from target), but could not be identified. In that case, the policy shock, if it is truly unrelated to contemporaneous economic conditions, could be interpreted as a 'policy experiment' to help identify the optimal policy.

The significance of (29) is that it highlights the fact that the rank condition for generic identification does not depend only on the dynamics of the targets, through condition (26), but also on the actual value of the structural parameters $(\theta, \rho, \sigma_\epsilon)$. Hence, the non-identification region is larger than would have been implied by (26) alone. This is in contrast to the standard IV regression model, of section II, where the concentration parameter is independent of the structural parameters.

**Weak identification** By continuity, we expect identification to be weak for all values of the parameters close to the non-identification region $\Omega_0$. It seems impossible to offer more precise remarks unless we consider specific cases.

For example, the effect of $\sigma_\epsilon^2$ and $\Sigma_{\eta\eta}$ on the concentration parameter $\mu_{min}^2$ is uncertain. From (25) we observe that $\Sigma_{vv}$ is unambiguously increasing in $\sigma_\epsilon^2$ and $\Sigma_{\eta\eta}$, thus affecting $\mu_{min}^2$ negatively. But $\sigma_\epsilon^2$ and $\Sigma_{\eta\eta}$ also affect $\mu_{min}^2$ positively through the signal in the first-stage regression, $\Sigma_{ZZ}^{\perp}$. So the overall effect is ambiguous.

Nevertheless, some limited understanding of the relationship between $\sigma_\epsilon$ and $\Sigma_{\eta\eta}$ and $\mu_{min}^2$ can be gained by looking at a specific example. The example we looked at contains a univariate target, inflation, say, such that $\Sigma_{\eta\eta} = \sigma_\eta^2$ and $\theta = \beta(1-\rho)$ are scalars. The transmission mechanism (22) is set to ARDL(1,1), with $a_1 = 0.9$ and $b_1 = -0.2$, implying inflation persistence as well as a negative effect of the lagged real interest rate. The structural parameters

are varied in the range $\rho \in [0, 0.9]$ and $\beta \in [1, 3]$. In this setting, $\mu^2$ is found to be strictly monotonically increasing in $\sigma_\epsilon^2$ and decreasing in $\sigma_\eta^2$ for all values of the structural parameters. That is, identification is stronger when the variance of the policy shock is higher and the variance of the inflation shock is lower.

Next, we observe that $\mu^2_{min}$ is decreasing in the maximal canonical correlation between $r_{t-1}$ and the optimal instruments. In the simple example of the previous paragraph, where $w_t = \pi_t$, substitution in (23) yields $\Pi' Z_t = \frac{a_1^2}{1-b_1\beta(1-\rho)}\pi_{t-1}$, so the only relevant instrument is $\pi_{t-1}$. The correlation between that and the exogenous regressor $r_{t-1}$ is then decreasing in the degree of smoothing $|\rho|$, i.e. as $\rho \to 0$ identification weakens. When $\rho = 0$ and $\sigma_\epsilon = 0$, $r_{t-1}$ and $\pi_{t-1}$ are perfectly collinear and, consequently, $\mu^2 = 0$.

### Identification analysis using a structural completing model

In the last section of their paper, Clarida, Galí, and Gertler (2000) use a fairly standard forward-looking business cycle model for inflation and the output gap to discuss the macro-economic implications of an interest rate rule like (19). Here, we comment on the implications of that business cycle model for the identification of the parameters of the interest rate rule.

The model consists of the equations:

$$\begin{array}{rcl} \pi_t &=& \delta\, \mathrm{E}(\pi_{t+1}|\mathcal{F}_t) + \lambda\, x_t \\ x_t &=& \mathrm{E}(x_{t+1}|\mathcal{F}_t) - \varphi^{-1}\,[r_t - \mathrm{E}(\pi_{t+1}|\mathcal{F}_t)] \end{array} \tag{30}$$

which, together with the interest rate rule equation (19) constitute a complete business cycle model for $y_t = (\pi_t, x_t, r_t)'$. This system can be written in the form (1):

$$B_0\, y_t = B_1\, \mathrm{E}(y_{t+1}|\mathcal{F}_t) + B_{-1}\, y_{t-1} + e_t \tag{31}$$

where the matrices $B_i$, $i = -1, 0, 1$ depend on the model's parameters ($\beta$, $\gamma$, $\rho$, $\delta$, $\lambda$, $\varphi$) and the vector of forcing variables $e_t$ contain the policy shock $\epsilon_t$ as well as any inflation and output shocks (which are omitted from (30)).

The existence and uniqueness of a non-explosive solution to this system depend on the roots of the polynomial $B(L) = \sum_{i=-1}^{1} B_{-i}\, L^i$. For existence, there must be at most 3 explosive roots. When this condition is satisfied with equality, and assuming $\mathrm{E}(e_t|\mathcal{F}_{t-1}) = 0$, the unique solution of the system (31) will be of the form:

$$y_t = C\, y_{t-1} + Q\, e_t \tag{32}$$

More specifically, the exclusion restrictions in (30) imply that $C\, y_{t-1} = c\, r_{t-1}$, for some given $3 \times 1$ vector of coefficients $c$. This corresponds to the case where $A_i = 0$ and $b_j = 0$ for all $i, j$ in (22), which is precisely one of the cases in which the Taylor rule parameters $(\beta, \gamma, \rho)$ are partially identified. In fact, since $\Pi = 0$ in (25), the entire concentration parameter is 0, not just its minimum eigenvalue, and hence the degree of under-identification is 2.

When the parameters are such that there are infinite backward solutions to (31), the conclusions about identification may be different. Clarida, Galí,

and Gertler (2000) map the space of the Taylor rule parameters $(\beta, \gamma, \rho)$ for which the solution is unique, for plausible values of the remaining parameters $(\delta, \lambda, \varphi)$. They find that $\beta > 1$ will always lead to a unique solution, while $\beta < 0.97$ will always lead to non-uniqueness and sunspot equilibria. In those cases, the presence of sunspot shocks can induce additional fluctuations in inflation and the output gap, beyond what is implied by fundamental shocks. Thus, rules with $\beta > 1$ are deemed stabilizing.

Therefore, we may conclude that if the above real business cycle framework is considered to be a reasonable approximation to reality, then stabilizing policy rules are expected to be weakly identified.

**Additional comments**

The main message of the analysis of the Taylor rule example was that identification problems arise when (but not exclusively) inflation and the output gap, or any linear combination of them, has very little dynamics. We acknowledge that such a situation may appear empirically implausible, in view of the large persistence evident in those macroeconomic time series. However, we wish to emphasize one important source of weak identification that may be empirically relevant.

Given some (possibly limited) knowledge of the transmission mechanism, the problem facing the policy maker can be decomposed in two steps. First, to align the interest rate with the predictable variation in inflation. Econometrically, that would be interpreted as making the interest rate correlate highly with the remaining determinants of inflation. This would weaken the identification of the Taylor rule (19) due to the collinearity between instruments and exogenous regressors that we discussed above.

The second step would be to make that correlation perfectly negative, so that the entire predictable variation in inflation disappears, and inflation becomes primarily driven by unanticipated shocks. In the words of a prominent central banker, active inflation targeting had precisely that effect:

> '[A]fter just a couple of years of [inflation] targeting, [...] expectations over a 2-year horizon [...] tended to be affected little by what was happening to current inflation rates. This was in marked contrast to earlier periods in Canadian history, in which expectations for the future had been fairly tightly linked to recently observed inflation rates.' (David Dodge, Governor of the Bank of Canada, Speech at the AEA annual meeting, Atlanta 2002).

In other words, the more successful the policy, the more inflation forecasts converge to the actual inflation target, and the less they depend on current and past data, which is a necessary condition for a forward-looking Taylor rule to be empirically identified.

To sum up, the source of weak identification in the forward-looking Taylor rule (19) lies in its theoretical underpinnings. When the monetary authority

is effective in controlling inflation, future inflation must correlate little with current economic conditions, and the bulk of its fluctuation must be due to (unpredictable) future shocks. This is precisely what leads to a low value for the concentration parameter. Thus we see that forward-looking policy rules will be least identified in periods when monetary policy has been most effective in controlling inflation. So, how can such equations be useful in providing reliable evidence that monetary policy has been effective in controlling inflation?

# IV Conclusion

In this paper, we analyzed the problem of weak identification of forward-looking models estimated with GMM, focusing on applications from the monetary economics literature. We discussed the various sources of weak identification, and a relevant measure with which to diagnose identification problems, the concentration parameter.

Our analysis showed that weak identification cannot be ruled out a priori for the estimation of either forward-looking Phillips curves or forward-looking monetary policy rules. Thus the existing empirical analyses of such models should be treated with caution. In the light of this criticism, it would be useful to re-evaluate the conclusions of the existing literature using inferential methods that are robust to weak identification, such as the conditional score and likelihood ratio tests proposed by Kleibergen (2002) and Moreira (2003).

# References


Anderson, T. W. (1977). Asymptotic expansions of the distributions of estimates in simultaneous equations for alternative parameter sequences. *Econometrica 45*, 509–518.

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

Batini, N. and A. G. Haldane (1999). Forward-looking rules for monetary policy. See Taylor (1999), pp. 157–192.

Batini, N., B. Jackson, and S. Nickell (2000). Inflaction dynamics and the labour share in the UK. Discussion paper 2, External MPC Unit, Bank of England, UK.

Binder, M. and M. H. Pesaran (1995). Multivariate rational expectations models: A review and some new results. In M. H. Pesaran and M. R. Wickens (Eds.), *Handbook of applied econometrics*, Volume Macroeconomics, pp. 139–187. Oxford: Blackwell.

Blanchard, O. J. and C. M. Kahn (1980). The solution of linear difference models under rational expectations. *Econometrica 48*, 1305–11.

Buiter, W. and I. Jewitt (1989). Staggered wage setting with real wage relativities: Variations on a theme of Taylor. In W. Buiter (Ed.), *Macroeconomic Theory and Stabilization Policy*, pp. 183–199. Ann Arbor: University of Michigan Press.

Choi, I. and P. 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.

Clarida, R., J. Galí, and M. Gertler (1998). Monetary policy rules in practice: Some international evidence. *European Economic Review 42*, 1033–67.

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.

Cragg, J. G. and S. G. Donald (1997). Inferring the rank of a matrix. *J. Econometrics 76*(1-2), 223–250.

Fuhrer, J. C. and G. R. Moore (1995). Inflation persistence. *Quarterly Journal of Economics 110*, 127–159.

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

Galí, J., M. Gertler, and J. D. López-Salido (2001). European inflation dynamics. *European Economic Review 45*, 1237–1270.

Hahn, J. and J. Hausman (2002). A new specification test for the validity of instrumental variables. *Econometrica 70*, 1517–1527.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*, 1029–1054.

Hansen, L. P., J. Heaton, and A. Yaron (1996). Finite sample properties of some alternative GMM estimators. *Journal of Business and Economic Statistics 14*, 262–280.

Hansen, L. P. and T. J. Sargent (1991). *Rational Expectations Econometrics*. Boulder: Westview Press.

Johansen, S. and K. Juselius (1994). Identification of the long-run and short-run structure. an application to the ISLM model. *Journal of Econometrics 63*, 7–36.

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica 70*(5), 1781–1803.

Kleibergen, F. and R. Paap (2003). Generalized reduced rank tests using the Singular Value Decomposition. Discussion paper 003/4, Tinbergen Institute, Amsterdam.

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. (2002). *Econometric Issues in Forward-Looking Monetary Models*. DPhil thesis, Oxford University, Oxford.

Mavroeidis, S. (2003). Identification and mis-specification issues in forward-looking models. Working paper, University of Amsterdam, Amsterdam.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica 71*(4), 1027–1048.

Pesaran, M. H. (1987). *The limits to Rational Expectations*. Oxford: Blackwell Publishers.

Phillips, P. C. B. (1984). Exact small sample theory in the simultaneous equations model. In S. Griliches and M. D. Ingriligator (Eds.), *The Handbook of Econometrics*, Volume 1, pp. 449–516. North-Holland.

Phillips, P. C. B. (1989). Partially identified econometric models. *Econometric Theory 5*, 181–240.

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

Rothenberg, T. J. (1984). Approximating the distributions of econometric estimators and test statistics. In S. Griliches and M. D. Ingriligator (Eds.), *The Handbook of Econometrics*, Volume 2, pp. 881–935. North-Holland.

Staiger, D. and J. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*, 557–586.

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

Stock, J. and M. Yogo (2003). Testing for weak instruments in linear IV regression. NBER technical working paper, 284, NBER, USA.

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

Taylor, J. B. (Ed.) (1999). *Monetary Policy Rules*. Chicago: University of Chicago Press.

Wright, J. H. (2003). Detecting lack of identification in GMM. *Econometric Theory 19*(2), 322–330.

# Appendix

### Coefficients in equation (17)

$$\alpha_1 = \frac{\lambda(\rho_1 + \beta\,\rho_2)}{1 - \beta\,\rho_1 - \beta^2\,\rho_2} \quad \text{and} \quad \alpha_2 = \frac{\lambda\,\rho_2}{1 - \beta\,\rho_1 - \beta^2\,\rho_2}.$$

**Derivation of (18)** Under stationarity, the variance of $s_{t-1}^{\perp}$ is derived from:

$$\text{var}(s_{t-1}^{\perp}) = \text{var}(s_t)\left[1 - \text{corr}(s_t, s_{t-1})^2\right] = \frac{1-\rho_2}{1+\rho_2}\frac{\sigma_\zeta^2}{(1-\rho_2)^2 - \rho_1^2}\left[1 - \frac{\rho_1^2}{(1-\rho_2)^2}\right].$$

**Derivation of equation (23)** Leading equation (22) one period and taking expectations conditional on $\mathcal{F}_t$, we have

$$\begin{aligned} w_{t+1|t} &= A_1 w_t + \sum_{i=1}^{n} A_{i+1}\, w_{t-i} + b_1 r_t + \sum_{j=1}^{m} b_{j+1}\, r_{t-j} \\ &= \sum_{i=1}^{n}(A_1 A_i + A_{i+1}) w_{t-i} + \sum_{j=1}^{m}(A_1 b_j + b_{j+1}) r_{t-j} + A_1\ \eta_t + b_1 r_t \\ &= \sum_{i=1}^{n}(A_1 A_i + A_{i+1}) w_{t-i} + \sum_{j=1}^{m}(A_1 b_j + b_{j+1}) r_{t-j} + A_1\ \eta_t \\ &\quad + b_1(\rho r_{t-1} + \theta' w_{t+1|t} + \epsilon_t). \end{aligned}$$

Hence

$$\begin{aligned} (I - b_1\theta') w_{t+1|t} &= \sum_{i=1}^{n}(A_1 A_i + A_{i+1}) w_{t-i} + \sum_{j=1}^{m}(A_1 b_j + b_{j+1}) r_{t-j} + \rho b_1 r_{t-1} \\ &\quad + A_1\eta_t + b_1\epsilon_t. \end{aligned}$$

We observe that $(I - b_1\theta')$ is invertible if and only if $\theta' b_1 \neq 0$. *Proof:* (if part) suppose $(I - b_1\theta')$ is singular s.t. $(I - b_1\theta')x = 0$, for some $x \in \Re^2$. Then $x = b_1(\theta' x) \in \mathrm{Col}(b_1)$, and $\theta' b_1 = 1$; (only if part) if $\theta' b_1 = 1$, then $\theta'(I - b_1\theta') = 0$.

Thus, when $\theta' b_1 \neq 1$, define $\delta = 1/(1 - \theta' b_1)$ and note that $(I - b_1\theta')^{-1} = I + \delta b_1\theta'$ and $(I - b_1\theta')^{-1} b_1 = \delta b_1$. So, the last equation simplifies to

$$\begin{aligned} w_{t+1|t} &= (I + \delta b_1\theta') \left( \sum_{i=1}^{n}(A_1 A_i + A_{i+1}) w_{t-i} + \sum_{j=1}^{m}(A_1 b_j + b_{j+1}) r_{t-j} \right) \\ &\quad + \delta b_1 \rho\ r_{t-1} + (I + \delta b_1\theta') A_1\ \eta_t + \delta b_1 \epsilon_t. \end{aligned} \tag{33}$$

Equation (23) then follows by substituting $w_{t+1} - \eta_{t+1}$ for $w_{t+1|t}$.

**The restricted reduced form** Noting that $\delta\theta' b_1 = \delta - 1$ and $\theta'(I + \delta b_1\theta') = (1 + \delta\theta' b_1)\theta' = \delta\theta'$, and substituting for $w_{t+1|t}$ from (33) into (20), yields the reduced form equation for $r_t$:

$$r_t = \delta\rho\ r_{t-1} + \sum_{i=1}^{n} \delta\theta'(A_1 A_i + A_{i+1}) w_{t-i} + \sum_{j=1}^{m} \delta\theta'(A_1 b_j + b_{j+1}) r_{t-j} + \delta\theta' A_1 \eta_t + \delta\epsilon_t.$$

Hence, the reduced form for the entire system $y_t = (w_t', r_t)' = (\pi_t, x_t, r_t)'$ is a VAR of order $k = \max(n, m)$, with reduced form residuals $u_t = (\eta_t', \eta_t' A_1'\theta\delta + \delta\epsilon_t)'$.