Downloaded from UvA-DARE, the institutional repository of the University of Amsterdam (UvA)
http://hdl.handle.net/11245/2.96144

File ID	uvapub:96144
Filename	467fulltext.pdf
Version	unknown

SOURCE (OR PART OF THE FOLLOWING SOURCE):

Type	pre-print - working paper
Title	Identification and mis-specification issues in forward-looking models
Author(s)	S. Mavroeidis
Faculty	FEB: Amsterdam School of Economics Research Institute (ASE-RI)
Year	2003

FULL BIBLIOGRAPHIC DETAILS:
http://hdl.handle.net/11245/1.346834

Copyright

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content licence (like Creative Commons).

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
(pagedate: 2014-11-21)

UvA ECONOMETRICS

Discussion Paper: 2002/21

Identification and mis-specification issues in forward-looking models

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

Identification and mis-specification issues in forward-looking models *

Sophocles Mavroeidis
University of Amsterdam

First version: August 2001
This version: March 2003

Abstract

Limited-information methods are commonly used to estimate forward-looking models with rational expectations, such as the "New Keynesian Phillips Curve" of Galí and Gertler (1999). In this paper, we address the issues of identification, which have been overlooked due to the incompleteness of the single-equation formulations. We show that problems of weak instruments can arise in these models, depending on the properties of the 'exogenous' variables, and that they are empirically relevant. We also uncover the link between identification and dynamic mis-specification, and examine the (lack of) power of the Hansen test to detect invalid over-identifying restrictions. Finally, with regards to the New Phillips curve, we conclude that this equation is either weakly identified or mis-specified, casting doubts on its utility as a model of inflation dynamics.

JEL classification: C22, E31

Keywords: Weak instruments, Rational Expectations, GMM, New Phillips Curve.

*I am grateful to David Hendry and Adrian Pagan for numerous comments and insightful discussions. I also benefited greatly from comments by Guillaume Chevillon, Kees Jan van Garderen, Eilev Jansen, Søren Johansen, Frank Kleibergen, George Konaris, John Muellbauer, Hugo Rodriguez, and seminar participants in Oxford, Amsterdam, the METU International Conference in Economics in Ankara (2001), the UK Econometric Study Group meeting in Bristol (2002), the 2nd workshop on international economics in Malaga (2002), and the Nordic Conference in Copenhagen (2003).

1 Introduction

Forward-looking Rational Expectations (RE) models are common in the macroeconomic literature. These models are typically of the form:

$$y_t = \beta \, \mathrm{E}(y_{t+1}|\mathcal{F}_t) + \gamma \, y_{t-1} + x_t \tag{1}$$

where y_t is a decision variable, x_t is a 'driving' or 'forcing' variable, usually thought of as 'exogenous', and $\mathrm{E}(y_{t+1}|\mathcal{F}_t)$ is the expectation of y_{t+1} conditional on the information set $\mathcal{F}_t$. The popularity of such models derives from the fact that the make the notion of forward-lookingness in economic decisions explicit and address the so-called Lucas (1976) critique.

The estimation of those models has been the subject of considerable research, see for instance Pesaran (1987) and Hansen and Sargent (1991) for some reviews. The various methods proposed in the literature can be divided into full and limited information methods, such as Full Information Maximum Likelihood (FIML) and Generalized Method of Moments (GMM), respectively. The former requires the specification of a model for the forcing variables x_t in addition to the structural equation (1), whereas the latter does not.

Traditionally, limited information methods have been more popular than the full information alternatives for a number of reasons. They appear to obviate the need to model the complete system (y_t, x_t), they are expected to work well under weaker conditions, and they involve relatively little efficiency loss against full information alternatives, whose implementation is typically much more demanding.[1]

[1] Full information methods require that the RE system be solved to derive the (restricted) reduced form or 'observable structure'. The nature of the solution, whether 'forward' (unique) or 'backward' (non-unique), requires a priori knowledge of the roots of the system, which depends on the actual values of the structural parameters as well as the parameters of the completing model for x_t. The resulting cross-equation restrictions, which are typically highly non-linear, can then be used to derive estimates of the structural parameters. An advantage of limited information methods, emphasized in Wickens (1993), is that they do not require the

However, there is a substantive condition that must be satisfied for any estimation method to provide consistent estimates and reliable inference on the parameters of interest β and γ. Namely, these parameters must be *identified* on the available information. The question of identifiability of such rational expectations models has been originally studied by Pesaran (1981). A thorough analysis of the order and rank conditions for the identifiability of rational expectations models, including (1), is given in Pesaran (1987). In that book, Pesaran warns against the "indiscriminate application of the IV method to RE models", stressing that the necessity of the identification condition "is often overlooked in the literature". Pesaran also urges that the conditions for identification of the RE model under consideration must be checked prior to estimation by limited information methods.

Unfortunately, this problem appears to have been overlooked in the recent monetary economics literature, where it has become common practice to estimate forward-looking rational expectations models by GMM. One example is the use of single-equation GMM for the estimation of forward looking monetary policy rules, popularized by Clarida, Galí, and Gertler (1998), see also Batini and Haldane (1999). 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, see also Batini, Jackson, and Nickell (2000) and Galí, Gertler, and López-Salido (2001).

These models provide the motivation for the present study. In particular, our analysis will focus on the New Keynesian Phillips curve model of Galí and Gertler (1999) (henceforth GG). It is clear that the same methodology can be applied to any forward-looking structural model.

However, this article is not intended merely as a reminder of the pitfalls of limited-

solution of the model prior to estimation.

information GMM estimation of under-identified forward-looking models. It makes a number of novel contributions.

Partial identification in this context has often been dismissed as a special case that does not merit particular attention.[2] However, this conclusion has recently been challenged by a growing literature which has come to be known under the name of 'weak instruments' or 'weak identification', see Stock, Wright, and Yogo (2002) for a review of the relevant theoretical and empirical contributions. Unlike under-identification, *weak* identification is not simply a special case, which is unlikely to arise in practice. It is highly relevant empirically, and it has been well-documented across the spectrum of applied econometrics. Thus, a main contribution of this paper is to demonstrate the empirical relevance of weak identification in forward-looking models and to alert researchers about the potential danger of weak instruments in monetary economics.

To this end, we propose (a variant of) the 'concentration parameter', originally introduced in the IV literature by Anderson (1977), as a measure of 'empirical' identification in forward-looking RE models. This measure, which has not been used in this context before, highlights that the strength of identification does not only depend on the dynamics of the forcing variables, as it can be deduced from Pesaran (1987, Proposition 6.2), but also on other features of the model that affect the degree of forecastability of future endogenous variables on past information.

Next, we emphasize the connection between identification and dynamic mis-specification of forward-looking models, which has implications for detecting lack of identification. Our results warn against relying solely on tests of rank reduction in the covariance between endogenous

[2] Wickens analyzes a similar model and asserts that "[l]ack of identifiability is not [...] a general feature", (Wickens 1993, p. 322).

regressors and the instruments, commonly advocated for structural equation models. We show that, in forward-looking models, such tests conflate identification and mis-specification, i.e. they have power also when the model is mis-specified.

Further, we discuss how identification of a ("pseudo-true") forward-looking rational expectations model can be achieved by imposing mis-specifying restrictions, e.g., in the form of omitted dynamics. We examine how detectable dynamic mis-specification is in that context, using the Hansen-Sargan test of overidentifying restrictions, which is the main mis-specification test under this methodology. The common practice in GMM estimation of forward-looking models has been to use many instruments, most of which are potentially irrelevant ('over-instrumenting'), and to robustify inference by means of general corrections for serial correlation ('over-correction'). Here, we expose the pitfalls of such an econometric practice.

Finally, all of this analysis is illustrated by applying it to the highly influential model of Galí and Gertler (1999), drawing some implications about its utility as an empirical model of inflation dynamics.

The structure of the paper is as follows. Section 2 introduces the hybrid Phillips curve model of GG and describes their baseline econometric formulation. Section 3 discusses the issues of partial identification and weak instruments with reference to that model. Section 4 addresses the issue of mis-specification, and studies its implications for the Hansen-Sargan test. Finally, section 5 concludes.

2 The model

'Phillips curve' is the name economists use to refer to an equation that describes the evolution of prices (or of the inflation rate) in a macroeconomic system. The New Keynesian Phillips curve is a pure forward-looking model of inflation dynamics, which typically takes the form:

$$\pi_t = \lambda\, x_t + \beta\, \mathrm{E}(\pi_{t+1}|\mathcal{F}_t) \tag{2}$$

where x_t is a forcing variable, usually a measure of the output gap or marginal costs. The set $\mathcal{F}_t$ contains, in principle, all of the information that is available to the agents at time t, which is usually more than a handful of macroeconomic variables that the econometrician may have at their disposal.[3]

Model (2) can be seen as a limiting case of a more general model that accommodates both forward- and backward-looking price-setting behaviour. Moreover, as it stands it is difficult to reconcile with the data. These considerations prompted a number of researchers to put forward a hybrid version of new and old Phillips curves, see, for example, Fuhrer and Moore (1995) and Buiter and Jewitt (1985):

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

GG proposed a new hybrid version, which is motivated by the idea of combining both

[3]Following the convention in the literature, e.g. Binder and Pesaran (1995), we will assume that $\mathcal{F}_t$ contains at least current and past values of the endogenous variable π_t, and the forcing variable x_t, namely $\mathcal{F}_t = (\pi_t, \pi_{t-1}, \ldots; x_t, x_{t-1}, \ldots; \ldots)$.

forward- and backward-looking price-setting behaviour. This leads to:[4,5]

$$\pi_t = \lambda\, s_t + \gamma_f \mathrm{E}(\pi_{t+1}|\mathcal{F}_t) + \gamma_b \pi_{t-1} + \epsilon_t \tag{4}$$

where $\epsilon_t \sim NID(0, \sigma_\epsilon^2)$ is adapted to $\mathcal{F}_t$, and it is an innovation with respect to $\mathcal{F}_{t-1}$, and s_t is a measure of real unit labour costs in deviations from their steady state, which is used as a proxy for marginal costs following Sbordone (2002).[6]

The above model cannot be estimated directly due to the fact that $\mathrm{E}(\pi_{t+1}|\mathcal{F}_t)$ is a latent variable. Therefore, we replace this expectation with the actual future realization in (4) in order to derive the GMM *estimating equation*:

$$\pi_t = \lambda s_t + \gamma_f \pi_{t+1} + \gamma_b \pi_{t-1} + e_t \tag{5}$$

where the 'structural residual' e_t is given by:

$$e_t = \epsilon_t - \gamma_f \eta_{t+1} \tag{6}$$

where $\eta_{t+1} \equiv \pi_{t+1} - \pi_{t+1|t}$ is the forecast error in predicting future inflation, and hence a mean innovation process with respect to $\mathcal{F}_t$.[7] This process can be explicitly derived, given a

[4]Their model also provides a link between what they call the "reduced form" parameters $(\lambda, \gamma_f, \gamma_b)$ in (4) and some "deep" or structural parameters (β, θ, ω). The latter have appealing interpretations as the discount factor, the degree of price inertia and the fraction of "backward-looking" agents, respectively. However, as pointed out in Mavroeidis (2002), the parametrization (β, θ, ω) is non-unique, and hence the latter are not globally identified, even when $(\lambda, \gamma_f, \gamma_b)$ are. This poses an additional difficulty for direct GMM estimation of (β, θ, ω), which we abstract from in this paper. For the implications of global but not local identification failure, see Ma (2002).

[5]This is a slightly generalized (and more realistic) version of the GG new hybrid Phillips curve model, see Galí and Gertler (1999, equation 26), where ϵ_t is absent. Whenever a *unique* solution to the rational expectations model exists (see section 3), that restriction, namely $\epsilon_t = 0$, would imply that the joint distribution of π_t and s_t is singular, with all the stochastic variation being driven by the latter. This is unnecessary for the validity of the model and need not be imposed. It would also imply that the structural residuals are serially uncorrelated, see (6), which is at odds with the data.

[6]In RE terminology, s_t, ϵ_t are the forcing variables, the latter being unobserved, and π_t is the endogenous or decision variable.

[7]$\pi_{t+1|t}$ is used synonymously with $\mathrm{E}(\pi_{t+1}|\mathcal{F}_t)$, by convention.

solution to the model. For example, define the 1-step forecast error $v_{t+1} = s_{t+1} - \mathrm{E}(s_{t+1}|\mathcal{F}_t)$. Then, under the solution (14) that we derive below:

$$\begin{aligned} \eta_{t+1} &\equiv \pi_{t+1} - \pi_{t+1|t} = \alpha s_{t+1} + \delta\pi_t + \frac{1}{1-\gamma_f\delta}\epsilon_{t+1} - \alpha\mathrm{E}(s_{t+1}|\mathcal{F}_t) - \delta\pi_t \\ &= \alpha v_{t+1} + \frac{1}{1-\gamma_f\delta}\epsilon_{t+1}. \end{aligned} \tag{7}$$

Hence, the structural residual e_t becomes

$$e_t = \epsilon_t - \gamma_f\alpha v_{t+1} - \frac{\gamma_f}{1-\gamma_f\delta}\epsilon_{t+1}.$$

We note that the e_t may exhibit negative serial correlation at lag 1 (if $\sigma_\epsilon^2 \neq 0$), without invalidating the model.

Equation (5) is an IV regression, with valid moment conditions of the form:

$$\mathrm{E}\left[(\pi_t - \lambda s_t - \gamma_f\pi_{t+1} - \gamma_b\pi_{t-1})Z_t\right] = 0 \tag{8}$$

for any $Z_t \in \mathcal{F}_t \backslash \{\pi_t\}$.[8] It appears that the parameters of interest $(\lambda, \gamma_f, \gamma_b)$ could be estimated by IV, or any other multi-step GMM procedure to take account of the serial correlation (and any potential heteroscedasticity) in the residuals, e_t. GG use a 2-step-2SLS estimator with a 12-lag Newey-West estimate of the covariance matrix of the moment conditions. However, any serial correlation beyond lag 1 is sufficient to invalidate their methodology, as we discuss in section 4.1 below.

3 Identification analysis

The identification analysis of the structural model requires knowledge of the reduced form of the system. This can be estimated directly from the data, or it can be derived from the

[8] π_t cannot be an admissible instrument for model (4), except in the special case $\sigma_\epsilon = 0$, since π_t is not orthogonal to ϵ_t, by construction.

structural model by postulating a distribution for the forcing variables and solving the system, as in Pesaran (1987). The latter approach assumes the structural model to be correctly specified, whereas the former does not.

Here, we use the latter approach for two reasons. First, simple structural equations such as the New Keynesian Phillips curve usually have considerable intuitive appeal, and this grants them some degree of immunity to criticisms of mis-specification. Nevertheless, even in such cases, it is important to examine whether the equation of interest is identified and hence empirically estimable, and whether inference based on it is reliable. Secondly, by assuming that the structural model is correctly specified, we can separate the analysis of identification from mis-specification issues, which are usually conflated. Thus, we can focus entirely on detecting pathological situations in which a correctly specified model becomes weakly identified.

For the sake of clarity, we will consider the following alternative distributions for s_t and discuss their implications for the identification of the parameters of interest, $(\lambda, \gamma_f, \gamma_b)$:

$$\text{Case 1}: \quad s_t = \rho s_{t-1} + v_t \tag{9}$$

$$\text{Case 2}: \quad s_t = \rho s_{t-1} + \varphi \pi_{t-1} + v_t \tag{10}$$

$$\text{Case 3}: \quad s_t = \rho_1 s_{t-1} + \rho_2 s_{t-2} + v_t \tag{11}$$

where $v_t \sim NID(0, \sigma_v^2)$ is an innovation w.r.t. $\mathcal{F}_{t-1}$ and it is orthogonal to ϵ_t in (4).[9] Also, $\mathcal{F}_t$ contains at least current and past values of the endogenous variable π_t, and the forcing variables (s_t, ϵ_t) (see footnote 3).[10]

[9] Normality is assumed for the purposes of simulation.

[10] GG use additional variables as instruments, as we discuss in section 3.3 below. However, whether this extra information is *relevant* (in predicting future inflation or the labour share) depends on the local DGP for (π_t, s_t), which has not been specified yet. Given the above 3 specifications of the completing process for s_t,

Case 1 is the benchmark under-identified case. It is a typical example of the simplest possible dynamic, stationary specification for an exogenous variable. The second and third cases are straightforward extensions of the first one, which introduce richer dynamics in the exogenous process. Case 2 introduces Granger causality of π_t for s_t, an extension that is not sufficient to yield identification.

In the third case, identification is achieved, but the 'strength' of the instruments crucially depends on the size of the second-order dynamics,[11] as well as on the variance ratio σ_ϵ/σ_v, see below. Also, since it nests the unidentified case, case 3 is suitable as a basis for discussion of weak instruments or lack of empirical identification. That is, we can study the behaviour of the estimators when the key identification parameter, ρ_2, becomes insignificant.

The completed model will consist of a pair of equations, one for inflation and the other for the conditioning variable, labour share. In other words, it will contain equation (4) together with one of (9) - (11).

3.1 Lack of identification

3.1.1 Case 1: $s_t \sim AR(1)$

The complete structural model under the leading case 1 would be:

$$\begin{aligned} \pi_t &= \lambda s_t + \gamma_f \mathrm{E}(\pi_{t+1}|\mathcal{F}_t) + \gamma_b \pi_{t-1} + \epsilon_t \\ s_t &= \rho s_{t-1} + v_t, \qquad \epsilon_t \perp v_t \end{aligned} \tag{12}$$

The total structural parameters are:

$$\boldsymbol{\theta} = (\lambda, \gamma_f, \gamma_b, \sigma_\epsilon^2, \rho, \sigma_v^2)$$

we will show below that the implied solution for inflation is such that no other variable should be relevant in predicting future inflation.

[11] This point goes back to Abel and Mishkin (1983), but appears to have been overlooked in the recent monetary economics literature. However, it is worth re-emphasizing it here.

of which $\boldsymbol{\theta}_1 = (\lambda, \gamma_f, \gamma_b, \sigma_\epsilon^2)$ are the *parameters of interest* and $\boldsymbol{\theta}_2 = (\rho, \sigma_v^2)$ are the *nuisance parameters.*

Assuming model-consistent expectations, the solution to this model will depend on the roots μ_i of the polynomial

$$\gamma(L) = 1 - \gamma_f L^{-1} - \gamma_b L \tag{13}$$

A necessary condition for the existence of a solution is that at most one of those roots is explosive, i.e., at least μ_1, say, is *not* explosive $|\mu_1| \geq 1$. When there is exactly one explosive root ($|\mu_2| < 1$), the solution is *unique.*[12] These conditions in turn impose restrictions on the set of admissible structural parameters $(\gamma_f, \gamma_b) \in (\Gamma_f, \Gamma_b)$. There are, of course, several different possibilities, but we shall only consider the range of values implicit in the estimates from GG, which are also plausible in many other contexts, namely:

$$\gamma_b, \gamma_f \geq 0 \quad \text{and} \quad \gamma_b + \gamma_f \leq 1$$

Both conditions are implied by the underlying model of GG.[13] These restrictions conveniently ensure (i) that the roots are real (and hence, rules out complex solutions), (ii) that a solution always exists, and, (iii) when the inequality is strict, that it is also unique.[14]

The possibility of non-uniqueness $\gamma_f + \gamma_b = 1$ is dealt in section 3.1.3 below. Hence, we

[12]The conditions for existence and uniqueness of solutions to rational expectations models are well-known, see Blanchard and Kahn (1980).

[13]This follows from the equation relating the 'deep' parameters (β, θ, ω) to $(\gamma_b, \gamma_f, \lambda)$, Galí and Gertler (1999, equation 25), and the requirement that the former lie between 0 and 1. These restrictions also hold in general for models of the form (4) that arise from a target-seeking decision problem under quadratic adjustment costs and uncertainty, see Pesaran (1987, example 7.2).

[14]The two roots are:

$$\mu_{1,2} = \frac{1 \pm \sqrt{1 - 4\gamma_f\gamma_b}}{2\gamma_b}$$

so that the discriminant $\Delta = 1 - 4\gamma_f\gamma_b \geq (1 - 2\gamma_b)^2$ is always positive. Hence $1 - 2\gamma_b + \sqrt{\Delta} \geq 1 - 2\gamma_b + \sqrt{(1 - 2\gamma_b)^2} = 0$ and $1 - 2\gamma_b - \sqrt{\Delta} \leq 1 - 2\gamma_b - \sqrt{(1 - 2\gamma_b)^2} = 0$, showing that $\mu_1 \geq 1$ and $\mu_2 \leq 1$. The last inequality is strict when $\gamma_f + \gamma_b < 1$, yielding a unique solution to the RE model.

can focus only on the unique solution to the structural model (12) which will be of the form:

$$\begin{aligned} &\pi_t = \alpha s_t + \delta \pi_{t-1} + u_t \\ &\alpha = \tfrac{\lambda}{1-\gamma_f(\delta+\rho)}, \qquad \delta = \tfrac{1-\sqrt{1-4\gamma_f\gamma_b}}{2\gamma_f}, \qquad u_t = \tfrac{1}{1-\gamma_f\delta}\epsilon_t \end{aligned} \tag{14}$$

with $\boldsymbol{\phi} = (\alpha, \delta, \sigma_u^2, \rho, \sigma_v^2)$ being the reduced-form parameters. The solution is the mapping from $\boldsymbol{\theta}$ to $\boldsymbol{\phi}$ and the latter characterize completely the DGP. Clearly, when no further restrictions are imposed on $\boldsymbol{\theta}$, it is of higher dimension than $\boldsymbol{\phi}$, so the mapping is not invertible and there is a multiplicity of $\boldsymbol{\theta}$s that correspond to the same DGP. Hence, the structural parameters cannot be jointly determined given knowledge of the reduced form, and the structural model (12) is *under*-identified. Since (ρ, σ_v^2) are in common, the under-identification affects $(\lambda, \gamma_f, \gamma_b, \sigma_\epsilon)$.

To see this differently, let us derive from the DGP (14) the conditional model for π_{t+1} given the variables that can be used as instruments, namely, $(s_t, s_{t-1}, \ldots, \pi_{t-1}, \pi_{t-2}, \ldots)$. This is essentially the first-stage regression of the endogenous regressor π_{t+1} on the instruments:

$$\pi_{t+1} = \alpha(\rho+\delta)\, s_t + \delta^2\, \pi_{t-1} + u_{t+1} + \alpha\, v_{t+1} + \delta\, u_t. \tag{15}$$

Now, subtract γ_f times (15) from the reduced form of inflation (14), and rearrange to get:

$$\pi_t = \gamma_f\, \pi_{t+1} + \underbrace{\alpha[1-(\delta+\rho)\gamma_f]}_{\lambda}\, s_t + \underbrace{\delta(1-\gamma_f\,\delta)}_{\gamma_b}\, \pi_{t-1} + \underbrace{(1-\delta\,\gamma_f)\, u_t - \gamma_f\, u_{t+1} - \gamma_f\, \alpha\, v_{t+1}}_{e_t}. \tag{16}$$

Viewing this as a GMM regression model with known (α, δ, ρ), clearly $(\lambda, \gamma_f, \gamma_b)$ are not jointly estimable from (16) with instruments $(s_t, s_{t-1}, \ldots, \pi_{t-1}, \pi_{t-2}, \ldots)$. Intuitively, the forward-looking parameter is un-identified because there are no relevant additional instruments in the forecasting regression (15) (beyond s_t and π_{t-1} which are already used as exogenous regressors) available to estimate it. Also, the other two parameters are functions involving the un-identifiable parameter γ_f, so they, too, will be un-identified. The degree of under-identification is only 1, but it spills over to all of the parameters of interest.

In general, to the extent that the reduced form model (solution) is nested within the structural model (as is the case here), there is trivially more than one observationally equivalent parameterization of the structural model.[15] Therefore, the issue here cannot be how to distinguish between a forward- and a backward-looking specification, which is impossible in this setting.[16]

3.1.2 Distributions of estimators under partial identification

The behaviour of GMM estimators under partial identification has been studied extensively in recent years. In linear models, with homoscedastic and uncorrelated errors Phillips (1989) and Choi and Phillips (1992) derived the asymptotic distribution of the IV estimator in partially identified models. They showed that the IV estimator of a completely un-identified parameter is $O_p(1)$ and its asymptotic distribution is 'Limiting Mixed Gaussian', centered on the probability limit of the respective least squares estimator. In contrast, the estimators of any identified parameters are $\sqrt{T}$-consistent, but their distribution is non-standard. When the parameters of interest are function of the unidentified parameters, their estimators will be inconsistent in general. These conclusions generalize to cases where the errors are heteroscedastic and/or autocorrelated, as well as to non-linear models estimated by GMM, see Stock and Wright (2000).

In addition, the distribution of GMM estimators and test statistics under partial identification depends on nuisance parameters which are not consistently estimable, such as the

[15] That is, setting $\gamma_f = 0$ in (12) yields (14), or simply, the solution to a backward-looking model is the model itself.

[16] This may become testable via tests for super-exogeneity, when we allow for the possibility of breaks in the parameters, see Hendry (1988) and Ericsson and Hendry (1999). We will not be concerned with this possibility here.

correlation of the endogenous regressors with the structural errors (a measure of 'endogeneity' of the regressors), or the structural error variance, see e.g. Nelson and Startz (1990), Staiger and Stock (1997). Conventional tests such as Wald, Likelihood Ratio or Score-type tests are not asymptotically similar and thus become unreliable, since their distribution can deviate arbitrarily from the assumed χ^2 asymptotic approximation, see Dufour (1997) and Wang and Zivot (1998).

When applied to the GG model under case 1, the above theoretical findings suggest that all of the GMM estimators of $(\lambda, \gamma_f, \gamma_b)$ should exhibit a 'double inconsistency', being both $O_p(1)$, and exhibiting a large bias in the direction of OLS. We investigate this by means of a Monte Carlo experiment.[17]

The Monte Carlo setting requires the specification of values for all the model's parameters. The parameters of interest are set to the values reported by Galí and Gertler (1999, Table 2), reproduced in table 1.[18] The first unrestricted parameterization is the one used for all the simulation experiments reported below. The restricted parameterization relates to the analysis of identifying restrictions that we discuss in this section. The remaining nuisance parameters $(\sigma_\epsilon, \rho, \sigma_v)$ are estimated from the GG data.

Table 1: Parameter values for simulation (i) and discussion of identifying restrictions (ii).

	ω	θ	β	γ_b	γ_f	λ	σ_ϵ	ρ	σ_v
(i)	0.486	0.834	0.909	0.378	0.591	0.015	0.18	0.9	0.10
(ii)	0.522	0.838	1.000	0.384	0.616	0.009	-	-	-

The different parameterizations are taken from Galí and Gertler (1999, table 2).

Next, we consider a 2-step GMM estimator, using the Newey and West (1987) Het-

[17]Some Monte Carlo evidence on the implication of IV estimation with "inadequate" instruments is also given in Wickens (1993).

[18]These estimates are based on a US GDP deflator-based measure of inflation, which is GG's preferred measure and the one we used for our data analysis later on.

eroscedasticity and Autocorrelation Consistent (HAC) estimator for the variance of the moment conditions (8).[19] The instrument set includes four lags of π_t and s_t, and a constant is included in the estimated equation. In order to replicate the GG analysis exactly, we are not using s_t as an instrument, which essentially means treating s_t as endogenous.[20]

We simulate both the GMM estimators and the respective OLS estimators of the parameters of interest $(\gamma_b, \gamma_f, \lambda)$. The results are given in table 2, and figure 1.

Table 2: Monte Carlo experiment of 2-step GMM - HAC(1) estimators of $(\lambda, \gamma_f, \gamma_b)$ in (12) under $s_t \sim AR(1)$.

	γ_b			γ_f			λ		
Sample	mean	st. dev.	bias	mean	st. dev.	bias	mean	st. dev.	bias
50	0.409	0.118	0.031	0.428	0.335	-0.163	0.053	0.273	0.038
100	0.423	0.102	0.045	0.426	0.319	-0.165	0.046	0.150	0.031
150	0.428	0.100	0.050	0.420	0.319	-0.171	0.045	0.113	0.030
300	0.430	0.094	0.052	0.423	0.302	-0.168	0.044	0.078	0.029
500	0.433	0.096	0.055	0.419	0.305	-0.172	0.045	0.069	0.030
1000	0.433	0.095	0.055	0.419	0.299	-0.172	0.044	0.060	0.029

The Monte Carlo Standard Errors (MCSE) are (decreasing in T)
γ_b: 0.0012 - 0.0009, γ_f: 0.0033 - 0.0030, λ: 0.0027 - 0.0006.
Parameter values for simulation: table 1, row (i).
plims of OLS estimators: $\gamma_{b,ols} = 0.43$, $\gamma_{f,ols} = 0.43$, $\lambda_{ols} = 0.042$.

The following observations stand out from the reported results. First, all the parameter estimates appear to be inconsistent, as anticipated. λ and γ_b are biased upwards relative to the 'true' values used for the simulation, whereas the forward coefficient γ_f is by far the

[19] This will be denoted HAC(i), where i refers to the lag truncation parameter.

[20] Strictly speaking, this is equivalent to estimating:

$$\pi_t = \lambda \mathrm{E}(s_t|\mathcal{F}_{t-1}) + \gamma_f \mathrm{E}(\pi_{t+1}|\mathcal{F}_{t-1}) + \gamma_b \pi_{t-1} + \varepsilon_t$$

where s_t is endogenous. This is justified by an 'error-in-variables' interpretation of s_t, when the latter is a *proxy* for the true relationship being driven by marginal costs, which is thus measured with 'error'. Note also that the above model is implied by (4), since it entails a weaker condition for the process of π_t. In other words, it is a weaker model that does not pin down the contemporaneous correlation between s_t and π_t.

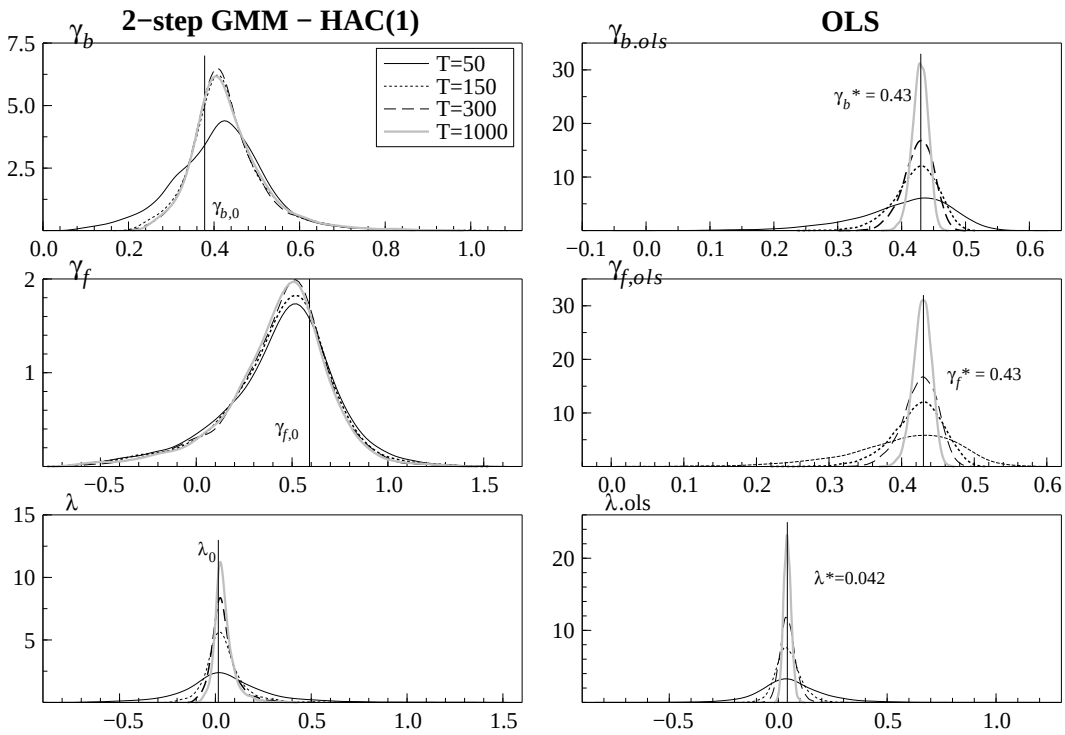

Figure 1: Simulated distributions of GMM and OLS estimators.

most affected parameter exhibiting a severe downward bias. More importantly, the coefficient estimators exhibit a bias that is almost identical to the OLS bias, in line with the asymptotic results outlined above.

Second, the variance of the estimator distributions falls very slowly with the sample size (compared to OLS), demonstrating that the estimators are of order $O_p(1)$ rather than $O_p(T^{-1/2})$. Even though the distribution of $\widehat{\lambda}$ seems to be imploding, the latter converges to the wrong value, the plim of the OLS estimator. Moreover, this implosion is misleading, since $\widehat{\lambda} = O_p(1)$ (indeed, increasing the sample $T > 1000$ doesn't reduce the variance of the estimator) and it simply indicates the prevalence of higher-order terms in small samples.[21]

[21] See Mavroeidis (2002).

3.1.3 Potentially identifying restrictions

Next, we discuss two theoretically motivated restrictions in terms of their implications for identification, and the extent to which they are testable.

The first restriction of particular interest is the following:

$$\mathrm{H}_\beta : \beta = 1 \quad \Leftrightarrow \quad \gamma_f + \gamma_b = 1 \tag{17}$$

Galí and Gertler emphasize this case because it links their model to the prototype hybrid Phillips curve of Buiter and Jewitt (1985). Also, since it reduces the number of structural parameters, we are interested in knowing whether it is sufficient for identification.

Imposing restriction (17) on (12) yields (letting $\gamma = \gamma_f$):

$$\begin{aligned} \pi_t &= \lambda s_t + \gamma \mathrm{E}(\pi_{t+1}|\mathcal{F}_t) + (1-\gamma)\pi_{t-1} + \epsilon_t \\ s_t &= \rho s_{t-1} + v_t. \end{aligned} \tag{18}$$

The analysis of section 3.1 showed that this restriction, together with $\gamma > 0$, guarantees existence of a solution of the form (14), but may also imply non-uniqueness. The lag polynomial (13) together with its roots now simplifies to:

$$\gamma(L) = 1 - \gamma L^{-1} - (1-\gamma)L, \qquad \text{with roots } \mu_1 = 1 \text{ and } \mu_2 = \frac{\gamma}{1-\gamma} \tag{19}$$

As far as uniqueness is concerned, we distinguish the following two interesting cases: $\gamma < 1/2$ and $\gamma \geq 1/2$.

Case $\gamma < \frac{1}{2}$: This implies $\mu_2 < 1$, and hence the solution is unique. The economic interpretation of this would be that the fraction of forward-looking agents-firms is smaller than the backward-looking ones. Or alternatively, partial adjustment (friction) dominates forward-looking behaviour.

The implied solution is:

$$\pi_t = \alpha s_t + \delta \pi_{t-1} + u_t \quad \text{with} \quad \alpha = \frac{\lambda}{1-\gamma-\gamma\rho}, \quad \delta = 1, \quad u_t = \frac{\epsilon_t}{1-\gamma} \tag{20}$$

Note that the coefficient on lagged inflation in the reduced form is 1, and independent of the structural parameters. This immediately implies that inflation has a unit root, which makes the restriction testable upon estimating the system (or the reduced form for inflation and performing a unit-root test). However, the structural model is still *under*-identified, since there are now 5 structural parameters $(\lambda, \gamma, \sigma_\epsilon, \rho, \sigma_v)$ and only 4 reduced form parameters $(\alpha, \sigma_u, \rho, \sigma_v)$.

Case $\gamma \geq \frac{1}{2}$: This is the case that best fits the GG data (see bottom line of table 1). This time the solution is non-unique, since none of the roots of the $\gamma(L)$ polynomial is explosive, and there is one non-predetermined variable. However, we can still solve the system forward using the unitary root of the (19) polynomial, since $|\rho| < 1$.[22] That solution would again be of the form (14)

$$\pi_t = \alpha s_t + \delta \pi_{t-1} + u_t \quad \text{with} \quad \alpha = \frac{\lambda}{\gamma(1-\rho)}, \quad \delta = \frac{1-\gamma}{\gamma}, \quad u_t = \frac{\epsilon_t}{\gamma} \tag{21}$$

Now the structural parameters become just-identified, although the strength of the identifying restriction remains an issue. To investigate it, consider following the re-parameterization of the structural equation (18):

$$\Delta \pi_t = \lambda s_t + \gamma \Delta_2 \pi_{t+1|t} + \epsilon_t \tag{22}$$

where $\Delta_2 = 1 - L^2$. Obviously, the strength of identification depends on the correlation between $\Delta_2 \pi_{t+1}$ and the extra instrument π_{t-1}. To study this, we derive the forecasting

[22] Note that the usual 'no bubbles' transversality condition is satisfied, by virtue of the fact that the resulting process for inflation will be stationary.

equation for $\Delta_2\pi_{t+1}$ given s_t and $(t-1)$-dated information, based on the solution (14):

$$\Delta_2\pi_{t+1|t} = \alpha(\rho+\delta)s_t + (\delta^2-1)\pi_{t-1}$$

Clearly, the identifying power of π_{t-1} for γ depends on δ. As δ gets closer to 1 (equivalently, γ gets closer to 1/2) identification weakens.

As we emphasized earlier, the under-identification of the unrestricted model implies that there is a *set* of values $\boldsymbol{\Theta}_0 \subset \boldsymbol{\Theta}$ for the structural parameters $\boldsymbol{\theta}$ that are consistent with the same reduced form (DGP), $\boldsymbol{\phi}_0$. Conceptually, only one of those values, say $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}_0$, will be 'true' in the sense that it corresponds to the true underlying behavioural relationship. However, we have no way of distinguishing between $\boldsymbol{\theta}_0$ and other values in $\boldsymbol{\Theta}_0$ in a statistical sense. Given that the degree of under-identification here is 1, imposing just one identifying restriction will yield identification (provided it is informative), and will thus ensure that the restricted estimators converge to a particular point $\boldsymbol{\theta}_0^R \in \boldsymbol{\Theta}_0$. However, in general, $\boldsymbol{\theta}_0^R$ need not correspond to the 'true' $\boldsymbol{\theta}_0$.

To illustrate the above point, let $\boldsymbol{\theta}_0 = (0.378, 0.591, 0.015)$ from the top line of table 1, and consider the restriction (17), $\gamma_f + \gamma_b = 1$.[23] The mean, standard deviation and mean bias of the restricted (2-step) estimators of (γ, λ) are estimated by simulation and are reported in table 3. The implied probability limits $(\boldsymbol{\theta}_0^R)$ of the restricted GMM estimators can be derived using equation (16) together with the restriction $\gamma_b = 1-\gamma_f = 1-\gamma$:

$$1-\gamma = \delta - \delta^2\gamma \quad \Rightarrow \quad \gamma = \frac{1}{1+\delta} \quad \text{and} \quad \lambda = \alpha\,[1-(\delta+\rho)\gamma].$$

Given the values $(\alpha_0 = 0.11, \delta_0 = 0.57, \rho_0 = 0.9)$ implied by substituting $\boldsymbol{\theta}_0$ into (14), we obtain $\gamma_0^R = 0.634$ and $\lambda_0^R = 0.007$.

[23] To confirm that this restriction is identifying, we derive the concentration parameter (see section 3.2) which is $1.94 \times T$, revealing strong identification.

The results reported in table 2 reveal convergence of the restricted GMM estimators to those implied true values. Since these are not the values that we used to simulate the data, the identifying restriction leads to inconsistency as we argued above. In this case, the imposition of the restriction tends to raise the forward-looking parameter (in line with the actual evidence reported by Galí and Gertler, compare the top and bottom lines in table 1).

Table 3: Monte Carlo experiment of GMM estimators of (λ, γ) in (18), under $s_t \sim AR(1)$.

	γ			λ		
Sample	mean	st. dev.	bias	mean	st. dev.	bias
50	0.609	0.093	0.018	0.012	0.150	-0.003
100	0.620	0.063	0.029	0.010	0.064	-0.005
150	0.624	0.050	0.033	0.008	0.043	-0.007
300	0.630	0.035	0.039	0.008	0.024	-0.007
500	0.632	0.027	0.041	0.008	0.016	-0.007
1000	0.635	0.019	0.044	0.007	0.010	-0.008
Parameter values for simulation: top line of table 1. Implied plims of estimators $\gamma_0^R = 0.634$, $\lambda_0^R = 0.007$. MCSE γ: 0.0013 - 0.0003, λ: 0.0021 - 0.0001.						

This was not an exhaustive account of what happens in the case $\gamma \geq 1/2$. As we pointed out at the beginning, the solution is non-unique and this indeterminacy gives rise to the possibility of effective over-identification.[24] However, this indeterminacy is particularly problematic, because the complete specification of the structural system (12) is insufficient to determine the local DGP (reduced form). Since different specifications of the reduced form have different implications for the identification of the structural model (4), the above identification analysis cannot be carried out without additional information, which could be obtained only by modelling the reduced form directly.

[24]See Mavroeidis (2002) for a detailed discussion of this case. This is an example of what Pesaran (1987) calls the 'irregular' case, discussed in section 6.4.2 of that book.

A second restriction of interest is the pure forward-looking specification:

$$\mathrm{H}_{\omega}: \ \omega = 0 \Rightarrow \gamma_b = 0$$

Provided $\gamma_f < 1$, the unique solution will involve setting $\delta = 0$ in (14), i.e.,[25]

$$\pi_t = \alpha s_t + u_t$$

Then, the number of structural parameters $(\lambda, \gamma_f, \sigma_\epsilon, \rho, \sigma_v)$ will still be in excess of the reduced form ones $(\alpha, \sigma_\epsilon, \rho, \sigma_v)$, leaving the system un-identified. However, note that this is only true under the assumption that the model is correctly specified, i.e., when the null hypothesis H_{ω} is correct. If we imposed $\gamma_b = 0$ incorrectly, i.e., when the correct model was (4) with unique solution (14), the restricted model would be just-identified. In that case, the implied estimate for the forward-looking coefficient can be derived from (16), by setting $\gamma_b = 0$, and it would be greater than 1.

3.1.4 Case 2: $s_t \sim ARDL(1,1)$

In the second leading case (10), where π_t Granger-causes s_t, the complete structural model becomes:

$$\begin{aligned} \pi_t &= \lambda s_t + \gamma_f \mathrm{E}(\pi_{t+1}|\mathcal{F}_t) + \gamma_b \pi_{t-1} + \epsilon_t \\ s_t &= \rho s_{t-1} + \varphi \pi_{t-1} + \upsilon_t \end{aligned} \tag{23}$$

[25] When $\gamma_f = 1$, we have the indeterminacy problem mentioned above. The model $\pi_t = \lambda\, s_t + \pi_{t+1|t} + \epsilon_t$ does not admit a unique solution, since both $\pi_t = \frac{\lambda}{1-\rho} s_t + \epsilon_t$ as well as $\Delta\pi_t = -\lambda\, s_{t-1} - \epsilon_{t-1} + \xi_t$ satisfy it, with ξ_t being any MDS w.r.t. $\mathcal{F}_{t-1}$.

This time, the solution of the model no longer depends on the roots of the polynomial (13), but rather on the roots of the matrix polynomial

$$\Gamma(L) = \begin{pmatrix} 1 & -\lambda \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \gamma_f & 0 \\ 0 & 0 \end{pmatrix} L^{-1} + \begin{pmatrix} \gamma_b & 0 \\ \varphi & \rho \end{pmatrix} L$$

So, the nuisance parameters (φ, ρ) come into play in determining the solution to the model. The possibility now arises that a solution may not exist (too many explosive roots, when φ is 'too big'). However, when φ is not too big, the conclusions of the simple AR(1) case hold, i.e., the model has a unique solution of the form (14) and hence the structural model remains un-identified. Thus adding a feedback from inflation to s_t is not sufficient to yield identification.

We also see another reason why s_t need not be weakly exogenous for the parameters of interest $(\lambda, \gamma_f, \gamma_b)$, in the sense of Engle, Hendry, and Richard (1983): the range of values these parameters can take is not independent of parameters of the marginal process (φ, ρ). In other words, the parameters of interest $(\lambda, \gamma_f, \gamma_b)$ are *not variation free* with respect to the parameter of the marginal process, if the system is to have a non-explosive solution. It seems that weak exogeneity of a forcing variable such as s_t, for the parameters of a forward-looking RE model, requires both Granger non-causality of the decision variable π_t for s_t, as well as sufficient dynamics in the forcing variable s_t.

In fact, by empirically estimating an ARDL(1,1) model for s_t on the GG data we notice that it is parsimoniously encompassed by the previous AR(1) specification (F(1, 149) = 1.33 with p-value 0.25). Also, the estimated feedback from lagged inflation is very small ($\widehat{\varphi} = 0.05$) and insignificant. Therefore, this case is not empirically relevant.

3.2 Case 3: a weakly identified system

We now proceed to the third leading case (11). This time the completed structural model is:

$$\begin{aligned} \pi_t &= \lambda s_t + \gamma_f \mathrm{E}(\pi_{t+1}|\mathcal{F}_t) + \gamma_b \pi_{t-1} + \epsilon_t \\ s_t &= \rho_1 s_{t-1} + \rho_2 s_{t-2} + v_t, \qquad \epsilon_t \perp v_t \end{aligned} \tag{24}$$

Since s_t again receives no feedback from lagged inflation, the solution to this model will be determined by the roots of the polynomial (13), and will be of the form:[26]

$$\pi_t = \alpha_1 s_t + \alpha_2 s_{t-1} + \delta \pi_{t-1} + u_t. \tag{25}$$

Now the number of reduced form parameters is the same as the number of structural ones, so the model is identified on the order condition. In fact, it is also identified on the rank condition when $\rho_2 \neq 0$, through the instrument s_{t-1}, as is verified by the forecasting equation for π_{t+1}:

$$\pi_{t+1} = (\alpha_1 \rho_1 + \alpha_2 + \delta \alpha_1) s_t + (\alpha_1 \rho_2 + \delta\, \alpha_2) s_{t-1} + \delta^2 \pi_{t-1} + \underbrace{u_{t+1} + \delta\, u_t + \alpha_1\, v_{t+1}}_{\zeta_t} \tag{26}$$

(contrast this with (15) in the previous un-identified cases).

The main issue is the strength of the identification. In line with the theoretical literature, this can be measured by the (eigenvalues of the) so-called *concentration parameter*, μ_T^2. This can be thought of as a multivariate signal-noise ratio in the first stage regression of the endogenous regressors on the instruments, and can be computed upon knowledge of the second moments in the data.[27] Using the parameter estimates from table 1 (i), the concentration

[26] δ and u_t are as in (14), whereas

$$\alpha_1 = \frac{\lambda(1-\delta\gamma_f)}{[1-\gamma_f(\delta+\rho_1)](1-\gamma_f\delta) - \rho_2\gamma_f^2}, \qquad \alpha_2 = \frac{\lambda\gamma_f\rho_2}{[1-\gamma_f(\delta+\rho_1)](1-\gamma_f\delta) - \rho_2\gamma_f^2}$$

which simplify to (14), when $\rho_2 = 0$.

[27] See, e.g. Stock, Wright, and Yogo (2002) for a simple definition, or Mavroeidis (2002), for the more general case of serially correlated and heteroscedastic structural residuals.

parameter is of the order of 10^{-3} for a sample size of 100. This is remarkably small, in the light of the results in the IV literature, where it is found that a concentration parameter with at least two digits is required for the conventional asymptotic theory to work.

3.2.1 Simulation results

We repeat the experiments of section 3.1, in order to compare the behaviour of the estimators in the two settings. We would also like to see how sensitive the results might be to the magnitude of the nuisance parameters, and how fast the asymptotic results are attained. The results are reported in table 4. The experiment in the top panel is based on the actual parameter values estimated from the data, and reveals weak identification, whereas the lower panel contrasts the results with an artificial situation of strong identification. Figure 2 compares the simulated distributions under weak and strong identification.

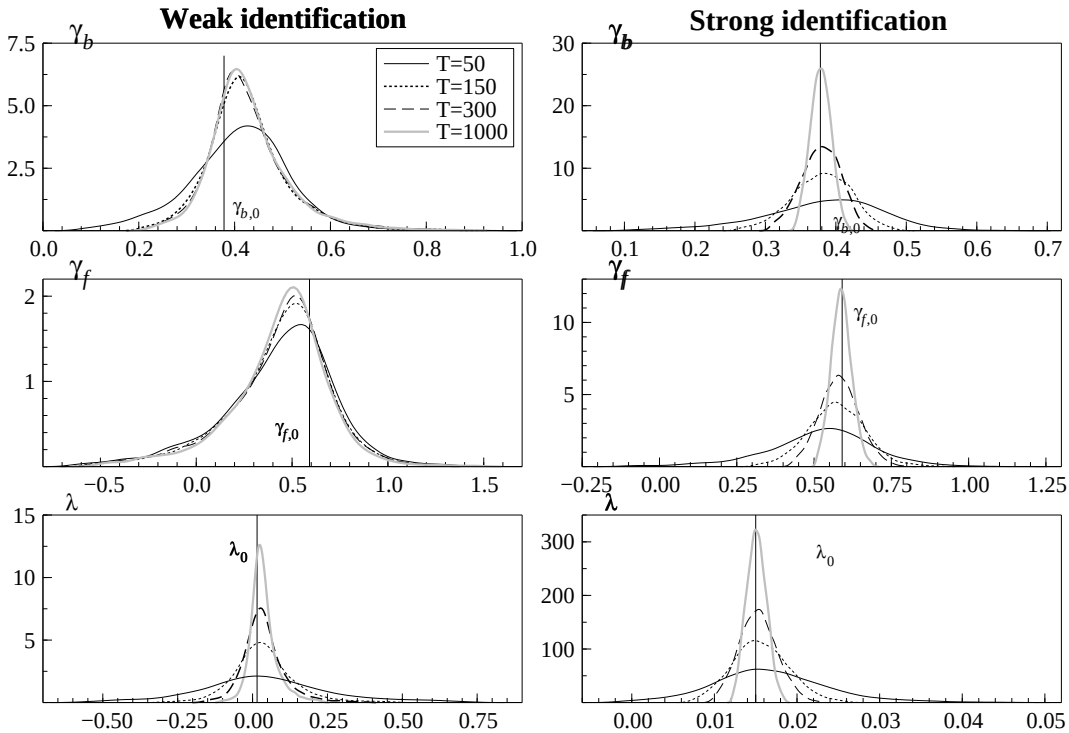

Figure 2: Simulated distributions of GMM estimators and weak and strong identification.

Table 4: Monte Carlo experiment of 2-step GMM – HAC(1) estimators of $(\lambda, \gamma_f, \gamma_b)$ in (24) under $s_t \sim AR(2)$.

(a) Weak identification: $\rho_1 = 0.9$, $\rho_2 = -0.05$, $\sigma_\epsilon = 0.18$, $\sigma_v = 0.1$.
Concentration parameter $\mu_T^2 = 10^{-5} \times T$.

	γ_b			γ_f			λ		
Sample	mean	st. dev.	bias	mean	st. dev.	bias	mean	st. dev.	bias
50	0.410	0.120	0.032	0.421	0.330	-0.170	0.044	0.297	0.029
100	0.421	0.099	0.043	0.429	0.322	-0.162	0.039	0.164	0.024
150	0.425	0.093	0.047	0.428	0.298	-0.163	0.038	0.119	0.023
300	0.428	0.094	0.050	0.426	0.305	-0.165	0.039	0.084	0.024
500	0.430	0.092	0.052	0.425	0.291	-0.166	0.037	0.063	0.022
1000	0.429	0.094	0.051	0.429	0.293	-0.162	0.036	0.052	0.021

MCSE: γ_b: 0.0017-0.0013, γ_f: 0.0047-0.0041, λ: 0.0042-0.0001

(b) Strong identification: $\rho_1 = 0.9$, $\rho_2 = -0.8$, $\sigma_\epsilon = 0.057$, $\sigma_v = 1$.
Concentration parameter $\mu_T^2 = 0.459 \times T$.

	γ_b			γ_f			λ		
Sample	mean	st. dev.	bias	mean	st. dev.	bias	mean	st. dev.	bias
50	0.378	0.094	0.000	0.513	0.206	-0.078	0.017	0.008	0.002
100	0.380	0.056	0.002	0.556	0.129	-0.035	0.016	0.005	0.001
150	0.379	0.044	0.001	0.570	0.103	-0.021	0.016	0.004	0.001
300	0.378	0.029	0.000	0.581	0.066	-0.010	0.015	0.002	0.000
500	0.378	0.022	0.000	0.586	0.049	-0.005	0.015	0.002	0.000
1000	0.378	0.015	0.000	0.588	0.034	-0.003	0.015	0.001	0.000

MCSE: γ_b: 0.0013-0.0002, γ_f: 0.0029-0.0005, λ: 0.0001-0.00001

For both panels: Values for structural parameters $(\gamma_b, \gamma_f, \lambda)$ taken from top line of table 1.
plims of OLS estimators $\gamma_{b,ols} = 0.43$, $\gamma_{f,ols} = 0.43$, $\lambda_{ols} = 0.042$.

It is not surprising to see that the baseline experiment results do not differ substantially from the previous unidentified case, given the weakness of the additional instrument s_{t-2}. Comparing table 4 (a) for case 3 with table 2 for the un-identified case 1, we observe only a small reduction in the mean bias of estimators and their standard deviation as a result of $\rho_2 \neq 0$. Similarly, comparing the left columns of figures 1 and 2 we see almost no difference in the shape of the estimator distributions.

Panel (b) of table 4 presents simulation results for a case in which the concentration parameter is high, giving rise to strong identification. To achieve a high value of the con-

centration parameter, we set $\rho_2 = -0.8$ (16 times bigger than the estimated value), and also make σ_v^2 100 times bigger and σ_ϵ^2 10 times smaller than their estimates from the data (none of these changes alone was sufficient). These values differ sharply from the empirical estimates based on the GG data.

The concentration parameter is a highly non-linear function of the nuisance parameters, whose exact analytical expression seems beyond reach. However, some intuition can be gained by thinking of it as a signal-noise ratio in the first stage regression (26), where the signal due to the instruments is $(\alpha_1\rho_2 + \delta\,\alpha_2)s_{t-1}$ and the noise is $\zeta_t = u_{t+1} + \delta\,u_t + \alpha_1\,v_{t+1}$. It can be shown that the variance of the signal is increasing in $|\rho_2|$, when ρ_1 is kept fixed. Also, since u_t is proportional to ϵ_t (see footnote 26), the variance of the noise is increasing in σ_ϵ^2. The effect of σ_v^2 is ambiguous, since it contributes both to the variance of the noise and to that of the signal, but apparently the latter effect dominates.

This discussion helps explain why identification based on the actual parameter estimates of $\rho_2 = -0.05, \sigma_v = 0.1$ and $\sigma_\epsilon = 0.18$ is weak (top panel of table 4). This is not only due to the second order dynamics in the forcing variable being weak (ρ_2 being small), but also because the standard error in the structural equation σ_ϵ is high relative to the variability of s_t.

This example also shows why it is dangerous to rely solely on the rank condition of identification, before proceeding with conventional estimation and inference. Even when the forcing variables have enough dynamics which would guarantee generic identification, e.g., a non-zero concentration parameter, the DGP may be such that the strength of identification remains a serious issue.

The analysis of this section showed that if we assume the Galí-Gertler model (4) to be *correctly specified*, and under a data coherent representation of the distribution of the forcing

variable s_t, the model is weakly identified on past information of π_t and s_t. Even at $T = 1000$ the estimators fail to converge to their true values. This situation is similar to other instances of weak identification, such as the Fuhrer, Moore, and Schuh (1995) linear-quadratic inventory model, where the authors report one case in which 30000 observations are needed for the GMM estimators to converge to the true value.

To corroborate this finding, we next conduct a formal test of identifiability.

3.3 Testing identifiability in the GG model

Partial identification of the structural model (4) corresponds to a rank reduction in the covariance matrix of the endogenous regressors with the instruments.[28] We test this using a Quasi-Minimum-Distance (QMD) statistic, which is a modification of the standard likelihood ratio (trace) test of Anderson and Rubin (1950) for reduced rank in the coefficients of the first-stage regression, to account for serial correlation and possible heteroscedasticity of the residuals. The former arises naturally in this model since the endogenous regressor π_{t+1} is projected on $t-1$ dated information.[29]

We give two versions of the test, one using only lags of s_t and π_t as instruments, and the other using additional variables that GG used. Those extra variables are commodity price and wage inflation, a measure of output gap and a long-short interest rate spread. Four lags of each of those variables are used as instruments. The results of the QMD test are given in table 5. Identification rank refers to the identifiability of the endogenous parameters, λ

[28] Which regressors are treated as endogenous by GG depends on their choice of instruments. Following GG, we exclude contemporaneous variables from the instrument set, hence leaving (π_{t+1}, s_t) as the endogenous variables in the model 5.

[29] The QMD test is developed in Mavroeidis (2002), and can be seen as a simple, reliable alternative to the more computationally demanding and potentially unstable procedures proposed in Cragg and Donald (1997).

Table 5: Identifiability Tests.

	Small Instrument Set, $k = 8$		Large Instrument Set, $k = 24$	
Rank (r)	QMD	LR	QMD	LR
0	418.31 [0.000]	227.44 [0.000]	612.17 [0.000]	365.80 [0.000]
1	23.55 [0.001]	22.046 [0.001]	144.50 [0.000]	129.24 [0.000]

The p-values are given in square brackets, and they are based on the χ^2 distribution with degrees of freedom $(k-1-r)(2-r)$.

and γ_f. Rank 0 corresponds to no identification, rank 1 to partial identification (a linear combination of $(\gamma_b, \gamma_f, \lambda)$ is identifiable) and finally rank 2 implies full identification. We also report the uncorrected (invalid) Likelihood Ratio reduced rank (LR) test for comparison.

We see that both reduced rank hypotheses 0 and 1 are strongly rejected in favour of complete identification. This is true for both the smaller and the larger instrument set. It is also evident that the additional instruments $(\text{gap}, dc, dw, spr)$ reinforce the identification of the structural parameters.

In the light of these tests, we see that the evidence on identifiability seems to conflict with the theoretical analysis. On the one hand, by looking at the assumed completing model for s_t we concluded that the structural model must be under-identified. On the other hand, the test on the strength of the empirical instruments reveals strong evidence in favour of identification. How can we reconcile these two contradictory findings?

The answer seems to lie in the fact that the original structural model may be mis-specified because of omitted dynamics, or other omitted variables, which are then incorrectly used as instruments. These possibilities are explored in the next section.[30]

[30] This contradiction could also be due to mis-specification of the completing process for s_t. However, it turns out that the AR(1) specification for s_t parsimoniously encompasses all models containing up to four lags of any of the other variables in the GG data set. Therefore, this is unlikely to be the source of the problem.

4 Mis-specification analysis

We will be concerned with one particular type of mis-specification, namely the violation of the moment restrictions. This may come about due to: either (a) lags of the dependent variable being incorrectly omitted from the structural equation and used as instruments; or (b) other variables in the information set being used as instruments incorrectly.

The first cause of mis-specification is likely to show up as higher-order serial correlation in the estimated structural residuals. The second need not have that side effect (although it might, if the omitted variables are themselves autocorrelated), and can be investigated by modelling the reduced form of the complete system.

This type of mis-specification admits two interpretations. It could be seen as a failure of rationality, when the model is assumed to be correctly specified; or as dynamic mis-specification of the model, i.e., the omission of some lagged exogenous variables, which are then used as instruments, incorrectly. If the latter interpretation is adopted, the model can be potentially extended to incorporate the extra dynamics, and satisfy the rational expectations condition. Indeed, that is exactly the approach followed in the literature when moving from the pure forward-looking Phillips curve (2) to a more general partial adjustment model (4). However, this flexibility comes at the cost of creating too many observationally equivalent models, as we show in the final section of the paper.

4.1 Omitted dynamics

To investigate these possibilities, we first examine the serial correlation pattern of the residuals from the GG model, noting that up to first-order is acceptable. Figure 3 displays the autocorrelogram and the partial autocorrelogram for the residuals from three different versions

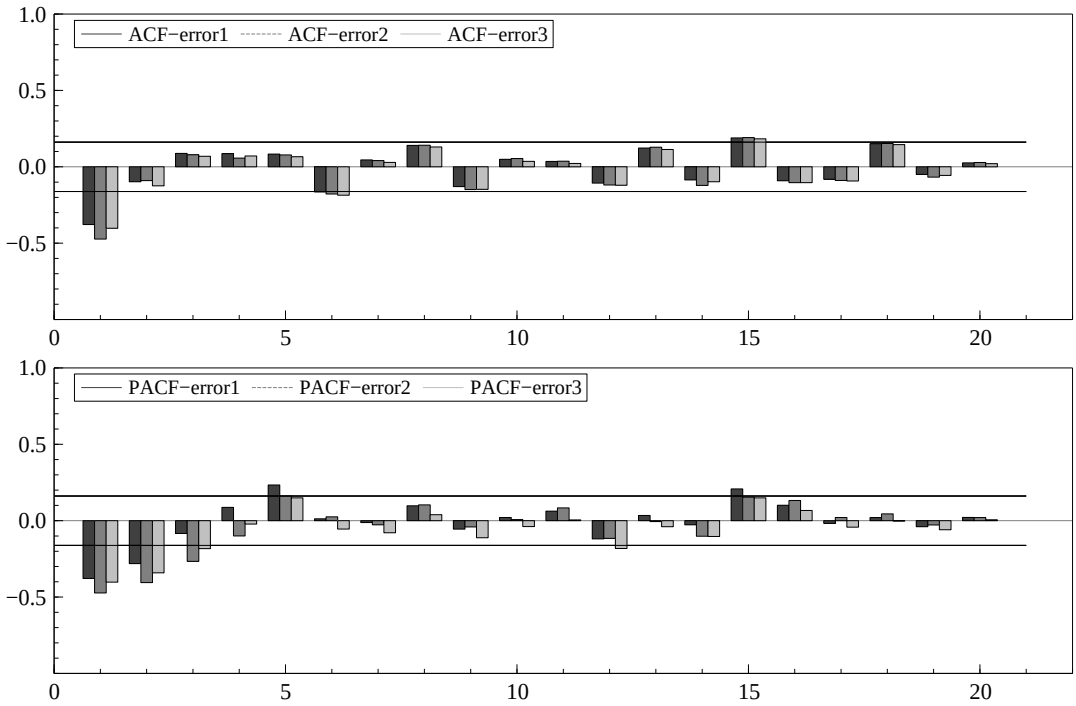

Figure 3: Autocorrelation of GG residuals.

of the GG model. The negative first-order autocorrelation is evident, and very significant. There is also some evidence of residual autocorrelation at further lags, albeit not very strong. Both the correlogram and the partial correlogram show that the MA(1) part of the series dominates.

The presence of any serial correlation beyond lag 1 may have two implications for the proposed model (4). If it is due to omitted dynamics from that model, it will immediately imply that the latter is mis-specified, and hence the resulting estimation and inference will be misleading. Otherwise, if the structural error ϵ_t in the model (4) is itself assumed to be autocorrelated, it will mean that some of the instruments used to estimate it may be invalid, i.e., those that lie within the autocorrelation horizon of the error. Suppose, for instance, that ϵ_t follows a MA(q) process. Then, π_{t-2} to π_{t-q} and possibly other variables up to lag q, would not be valid instruments.

In this analysis, we will focus on the first interpretation of serial correlation which is symptomatic of mis-specification. The reason is that the other interpretation could be dealt with in this framework by selecting the instruments appropriately. Yet, it has to be emphasized that it is methodologically incorrect to assume qth order serial correlation in the structural errors ϵ_t, say, and still use lags 1 to q of the dependent variable(s) as instruments.[31] Moreover, even if such an interpretation is given to the observed serial correlation and the instrument set is adapted accordingly (removing the appropriate lags of y_t), then the "excess serial correlation" which is symptomatic of mis-specification, will refer to the serial correlation beyond what is implicitly allowed for in the selection of instruments.

To analyse the implications of excess serial correlation, we need to generalize the DGP, so that the structural model (4) becomes dynamically mis-specified relative to it. Given the specificity of Monte Carlo experiments, rather than arbitrarily introducing further dynamics in the DGP, we estimate them from the data. Namely, we model the reduced form as a parsimonious ARDL(p,q) for π_t given s_t and then 'invert' this to find a forward-looking specification that has this reduced form as its solution.

The estimated reduced-form equation for inflation on the GG data is

$$\widehat{\pi}_t = \underset{(0.0631)}{0.68}\ \pi_{t-1} + \underset{(0.0617)}{0.241}\ \pi_{t-3} + \underset{(0.161)}{0.554}\ s_t - \underset{(0.162)}{0.476}\ s_{t-2} \tag{27}$$

with $\widehat{\sigma}_u = 0.26$ and standard errors given in the parentheses (the omitted lags were insignificant). Symbolically, the reduced form can be written as:

$$\pi_t = \alpha\, s_t + \alpha_2\, s_{t-2} + \delta_1\, \pi_{t-1} + \delta_3\, \pi_{t-3} + u_t. \tag{28}$$

[31] In particular, if the serial correlation of the structural errors is assumed to be potentially infinite, *no* lag of the dependent variable could be used as an instrument.

Equation (28) together with an AR(1) specification for s_t enable us to derive $\mathrm{E}(\pi_{t+1}|\mathcal{F}_t)$:

$$\pi_{t+1|t} = \alpha\,\rho\,s_t + \alpha_2 s_{t-1} + \delta_1\pi_t + \delta_3\pi_{t-2}. \tag{29}$$

Given some value for the forward-looking parameter γ_f, we can subtract $\frac{\gamma_f}{1-\gamma_f\delta_1}$ times (29) from both sides of equation (28) to get an isomorphic forward-looking specification:

$$\pi_t = \lambda s_t + \lambda_1 s_{t-1} + \lambda_2 s_{t-2} + \gamma_f \mathrm{E}(\pi_{t+1}|\mathcal{F}_t) + \gamma_1\pi_{t-1} + \gamma_2\pi_{t-2} + \gamma_3\pi_{t-3} + \epsilon_t \tag{30}$$

where the structural parameters $(\lambda, \lambda_1, \lambda_2, \gamma_1, \gamma_2, \gamma_3)$ are functions of the reduced form parameters.[32] When γ_f is set such that the lag polynomial $1 - \gamma_f L^{-1} - \gamma_1 L - \gamma_2 L^2 - \gamma_3 L^3$ has exactly one explosive root, (28) is the unique solution to the rational expectations model (30).

Obviously, the generalized forward-looking specification (30), which is consistent with the GG data, implies that the baseline equation (4) is dynamically mis-specified due to the omission of $(\pi_{t-2}, \pi_{t-3}, s_{t-1}, s_{t-2})$. Moreover, the structural equation (30) is unidentified, since there is an infinity of observationally equivalent models of the form (30) for different arbitrary choices of γ_f. On the other hand, the restricted equation (4) is now well-identified through the instruments $(\pi_{t-2}, \pi_{t-3}, s_{t-1}, s_{t-2})$, although this identification is achieved via mis-specification.

The question that naturally arises is: how detectable is that mis-specification through a test of over-identifying restrictions? We address this question by means of a small Monte Carlo experiment on the power of the Hansen-Sargan test with particular emphasis on the implications of 'over-instrumenting' and 'over-correction' for serial correlation.

[32] Matching coefficients yields: $\lambda = \alpha[1 - \gamma_f(\delta_1 + \rho)]$, $\lambda_1 = -\alpha_2\,\gamma_f$, $\lambda_2 = \alpha_2(1 - \gamma_f\delta_1)$, $\gamma_1 = \delta_1(1 - \gamma_f\delta_1)$, $\gamma_2 = -\delta_3\,\gamma_f$, $\gamma_3 = \delta_3(1 - \gamma_f\delta_1)$ and $\epsilon_t = (1 - \gamma_f\delta_1)\,u_t$.

4.2 The Hansen-Sargan test

The Hansen-Sargan test (also referred to as the J-test) is well-known to have a $\chi^2(k-p)$ distribution asymptotically under the null, with degrees of freedom equal to the degree of over-identification $k-p$. In the presence of mis-specification, the distribution of the statistic can be approximated by a non-central $\chi^2(k-p,\, \nu^2)$, in large samples. Its non-centrality parameter, ν^2, which can be computed exactly when the DGP is known, will be referred to as the 'mis-specification parameter', since it measures the extent to which mis-specification is detectable.[33] Therefore, the power of the J-test can be characterized approximately by ν^2 and its degrees of freedom.

We consider four different versions of the J-test using different numbers of instruments (k= 8 and 24);[34] and two different types of serial correlation corrections: HAC(1) refers to the parametric West (1997) MA-l estimator allowing only up to first-order residual autocorrelation, and HAC(12) refers to the nonparametric Newey and West (1987) estimator with lag-truncation parameter 12. The DGP is given by equations (27) and (9) and the estimated model is (5).

Table 6 reports the rejection frequencies of the test statistic at the 5% level, under this fixed alternative. Notably, the concentration and mis-specification parameters are increasing (linearly) with the sample size, reflecting stronger identification and mis-specification of the model.

Several conclusions stand out from those results. First, we see that the identification and mis-specification issues in this model are conflated. This is generally true in forward-looking

[33]Similarly to the concentration parameter, which measures the strength of identification.

[34]The small set contains only four lags of the variables in the model s_t and π_t, whereas the larger set resembles the instruments that GG used.

Table 6: Rejection frequencies for the Hansen-Sargan test of over-identifying restrictions, 5% level (3 estimated parameters and k instruments).

			k= 8			k=24		
T	μ_T^2	ν_T^2	HAC(12)	HAC(1)	$\chi^2(5, \nu_T^2)$	HAC(12)	HAC(1)	$\chi^2(21, \nu_T^2)$
50	4.5	1.444	0.000	0.091	0.124	0.000	0.020	0.081
150	13.5	4.333	0.000	0.349	0.315	0.000	0.142	0.166
300	27.0	8.667	0.308	0.637	0.604	0.000	0.411	0.335
1000	90.0	28.888	0.988	0.994	0.994	0.602	0.947	0.928

The DGP consists of equations (27) and (9) with $\rho = 0.9$ and $\sigma_v = 0.1$, and the estimated model is equation (5). 10^5 Monte Carlo replications used (MCSE = 0.0007).

models when leads of the regressand appear on the right hand side. This is because any variables that are incorrectly omitted from the structural model (compare (4) with (30)) will also serve as relevant instruments in predicting the endogenous regressors. In a sense, the endogenous lead regressor 'mops up' part of the mis-specification.[35] In the case of the GG model, we see that the omitted dynamics contribute more towards the identification than towards the mis-specifiability of the model.

Secondly, the addition of a large number of irrelevant instruments reduces the rejection frequencies of the test, as anticipated. Intuitively, the same degree of mis-specification is diluted over many instruments, thus reducing the test's power to detect it. In other words, 'over-instrumenting' unambiguously reduces power.

Thirdly, we notice a marked difference in the behaviour of the test using different autocorrelation corrections. It is well-known that the J-test statistic is non-pivotal with respect to the concentration parameter under weak instruments, see e.g. Stock and Wright (1997). Thus, we would expect some deviation from the asymptotic $\chi^2(k-p, \nu_T^2)$ distribution for low values of μ_T^2. However, the HAC(1)-based test exhibits relatively small deviations from its asymptotic power, even at small samples. In contrast, the HAC(12)-based version suffers a

[35] A similar argument was made independently by Rudd and Whelan (2001) with reference to the GG model.

severe downward bias in all cases. This is in line with extensive Monte Carlo evidence on the small sample properties of the various HAC estimators, and points out the potentially serious costs of 'over-correction' for serial correlation.[36]

The above results help explain why GG failed to detect any mis-specification with the J-test using 24 instruments and correcting for 12th-order autocorrelation. The p-value of this test for the GG data is 0.97, well within the acceptance region of no mis-specification.[37]

4.3 Anything goes...

In this final section we show how the un-detectability of mis-specification of structural models can give rise to several, apparently over-identified but almost observationally equivalent models, with significantly different estimates for the parameters of interest.

The most important parameter in this model is arguably the coefficient on the inflation lead in the structural equation, γ_f. This is interpreted as a measure of 'forward-lookingness' in the determination of inflation, as we discussed above. Table 7 reports the different point estimates and associated t-values obtained for γ_f using various generalizations of the baseline structural equation (4), in the spirit of the analysis in the previous section. In the third column we give a list of the variables that are included as exogenous regressors in the estimated structural equation. It is assumed that all remaining variables in the instrument set should

[36] See, e.g., den Haan and Levin (1997) and Mavroeidis (2002). A partial explanation of the poor performance of the HAC(12)-based test arises from the fact that the nonparametric HAC(12) estimator converges at a slower rate to the true asymptotic variance of the moment conditions than the parametric HAC(1), since HAC(12) $= O_p(T^{-1/3})$ versus HAC(1) $= O_p(T^{-1/2})$, see West (1997).

[37] Another example comes from the estimation of forward-looking monetary policy rules by Clarida, Galí, and Gertler (1998), where the authors estimate the same model for six different industrialized countries and report p-values for the J-test of 0.999 in all cases. Even though the different data sets may be correlated, the likelihood of this event happening under the null of no mis-specification is implausibly small when the test is correctly sized, suggesting that it may actually be biased in that case too.

be used as instruments. The last two columns give the results of identification (QMD) and mis-specification (J) tests. These results are based on a 2-step GMM estimator allowing only for first-order residual serial correlation (HAC(1)).

Table 7: Alternative estimates of γ_f using different models.

$\widehat{\gamma}_f$	t-prob	Exogenous regressors	Identif.	Mis-spec.
0.19	[0.244]	$\pi_{t-1}, \pi_{t-4}, s_{t-1} \ldots s_{t-3}, \text{gap}_{t-1} \ldots \text{gap}_{t-3}$, $dw_{t-2}, dw_{t-3}, dc_{t-1} \ldots dc_{t-4}$	[0.002]	[0.764]
0.25	[0.047]	$\pi_{t-1}, \pi_{t-4}, s_{t-1} \ldots s_{t-3}, \text{gap}_{t-1} \ldots \text{gap}_{t-3}$, $dw_{t-2}, dw_{t-3}, dc_{t-2}, dc_{t-4}$	[0.087]	[0.823]
0.31	[0.005]	$\pi_{t-1}, \pi_{t-4}, s_{t-1}, s_{t-2}, \text{gap}_{t-1} \ldots \text{gap}_{t-3}$, $dw_{t-2}, dw_{t-3}, dc_{t-2}, dc_{t-4}$	[0.000]	[0.889]
0.41	[0.000]	$\pi_{t-1}, \pi_{t-4}, s_{t-1}, s_{t-2}, \text{gap}_{t-1}, \text{gap}_{t-2}, dw_{t-3}, dc_{t-4}$	[0.000]	[0.771]
0.56	[0.000]	$\pi_{t-1}, \pi_{t-4}, dw_{t-3}, dc_{t-4}$	[0.000]	[0.860]
0.61	[0.000]	π_{t-1}	[0.000]	[0.884]

As it is shown in the table, the point estimate and inference on the parameter of interest is very sensitive to the way we choose to use our information, i.e. the choice of identifying restrictions. Also, in none of those models does the J-test reject the validity of the instruments, which are also found to be strongly identifying.

The above discussion demonstrated the dangers of 'over-instrumenting' and 'over-corrections' in the estimation of forward-looking models. When mis-specification is difficult to detect by standard tests of over-identifying restrictions, the estimators may become biased in unknown directions. In that case, structural models of this form may almost become just-identifying re-parameterizations of the reduced form. That in itself is not a criticism of structural econometric modelling. Indeed, such re-parameterizations are desirable whenever they add useful economic intuition to a model. However, the use of invalid inferential procedures to provide evidence in favour of a just-identified model seems to be methodologically unjustifiable.

5 Conclusions

In this study, we have been concerned with the problem of weak or partial identification of the parameters of a single equation structural econometric model with forward-looking rational expectations. In particular, in the first part we presented an economic-theoretic identification analysis, and compared it with a statistical testing procedure. In the second part, we discussed how identification may be achieved through mis-specification. In this setting, we examined the power of the standard mis-specification J-test when too many instruments and general autocorrelation corrections are used. Our conclusions point to potential dangers of this type of econometric practice.

First, when the model is weakly identified, the use of apparently valid moment restrictions may obscure the poor identification of the parameters. Namely, adding irrelevant instruments on the basis of rational expectations restrictions, say, would result in an apparent 'over-identification' when the model may in fact be un-identified. Weak identification will lead to inconsistent estimates of the parameters of interest, and the addition of extra instruments will merely induce bias in the direction of OLS estimates.

Second, modelling the 'exogenous' driving process in the rational expectation model may prove highly informative in uncovering pathological situations in which the parameters of interest are poorly identified. In the case of the New-Keynesian Phillips curve, these pathological situations appear to be relevant empirically. However, identification could be achieved indirectly through mis-specification.

Third, when the model is mis-specified due to omitted dynamics, the use of too many lagged instruments as well as too general autocorrelation corrections is likely to reduce the power of mis-specification tests, and obscure the inadequacies of the structural model. In that

sense, looking at the reduced form of the complete system may prove a valuable alternative to the single-equation approach, as it may help uncover that mis-specification.

Forth, the presence of undetectable mis-specification may result in a multiplicity of almost observationally equivalent models, giving rise to the possibility of choosing a desirable model within this class. As a result, tests of over-identification lose their strength as evidence in favor of the proposed structure, (and therefore as evidence of forward-looking behaviour).

In practice, careful selection of instruments may help avoid the problem of over-instrumenting that we analyzed here. Instead of that, a popular alternative is to include a large number of potentially relevant instruments, so as to maximize the chances of getting identification. We hope that the evidence presented here will deter researchers from following that route.

Finally, with regards to the estimation of the New Phillips curve of Galí and Gertler (1999), our analysis reveals that it is either weakly identified, or more probably mis-specified, casting doubts about its utility as a model of inflation dynamics.

A Supplementary notes

The concentration parameter

Consider the prototype linear IV model:

$$y_t = \theta' Y_t + u_t \tag{31}$$

$$Y_t = \Pi' Z_t + v_t, \tag{32}$$

where Y_t is a p-vector of endogenous regressors, Z_t is a k-vector of instruments, $u_t \sim iid(0, \sigma_u^2)$, and for a sample of random vectors $\{a_t, b_t\}$ define the notation $\Sigma_{ab} = \underset{T\to\infty}{\text{plim}} T^{-1} \sum_{t=1}^{T} a_t b_t'$. The concentration parameter is a matrix of dimension p which is given by:[38]

$$\boldsymbol{\mu}_T = T\, \Sigma_{vv}^{-1/2} \Pi'\, \Sigma_{ZZ}\, \Pi \Sigma_{vv}^{-1/2}. \tag{33}$$

When the structural error u_t is serially correlated and/or heteroscedastic (which is the case for the model analyzed here), a generalization of the above quantity is

$$\boldsymbol{\mu}_T = T\, \sigma_u^2\, \Sigma_{vv}^{-1/2} \Pi'\, \Sigma_{ZZ}\, V(\theta_0)^{-1}\, \Sigma_{ZZ}\, \Pi \Sigma_{vv}^{-1/2} \tag{34}$$

where $V(\theta_0) = A\text{var}\left(T^{-1/2} \sum Z_t\, u_t\right)$ is the asymptotic variance matrix of the moment conditions, and the scaling factor σ_u^2 is applied to correct for the units of measurement in y_t (see Mavroeidis 2002 for a justification). Note that in the special case of no heterescedasticity or serial correlation, (34) reduces to (33).

As argued by several authors, e.g., Stock and Yogo (2003), the minimum eigenvalue of the concentration parameter can serve as an index of identification. In this paper, the structural equation of interest contains two endogenous regressors (π_t, s_t), and the value of the concentration parameter reported in the text is the *minimum eigenvalue* of (34). This depends on the reduced-form parameters $(\alpha_1, \alpha_2, \delta, \rho_1, \rho_2, \sigma_u, \sigma_v)$, which characterize the local DGP, and hence the second moments of the

[38] When the instruments are strongly exogenous, and interest centers on the distribution of $\widehat{\theta}_{IV}$ *conditional* on Z_t, $T\,\Sigma_{ZZ}$ can be replaced by $Z'Z$. Also, when the structural equation (31) contains exogenous regressors, the formula is corrected for those regressors (e.g., re-interpret Y_t and Z_t as being corrected for the exogenous regressors X_t, by orthogonal projection).

data. The latter, in turn, depend on the structural parameters $(\lambda, \gamma_f, \gamma_b)$ and the nuisance parameters $(\rho_1, \rho_2, \sigma_v, \sigma_\epsilon)$. Fixing $(\lambda, \gamma_f, \gamma_b)$ and ρ_1 to the values reported in table 1, we may write μ_T^2 as:

$$\mu_T^2 \equiv \underset{\lambda}{\operatorname{argmin}} \left(|T\, \sigma_u^2\, \Pi'\, \Sigma_{ZZ}\, V(\theta_0)^{-1}\, \Sigma_{ZZ}\, \Pi - \lambda\, \Sigma_{vv}| \right) = T\, f(\rho_2, \frac{\sigma_v}{\sigma_\epsilon}, \sigma_\epsilon) \tag{35}$$

The above expression is a highly non-linear function of the nuisance parameters, whose exact analytical expression seems beyond reach. This implicit formulation emphasizes the dependence of μ_T^2 on the nuisance parameters $(\rho_2, \sigma_\epsilon, \sigma_v)$ and its linearity in T. It can also be shown that $f()$ is increasing in $|\rho_2|$ and σ_v, decreasing in σ_ϵ, and that it is homogeneous in $(\sigma_v, \sigma_\epsilon)$. The following table illustrates:

μ_T^2 for various values of ρ_2 and σ_v/σ_ϵ							
ρ_2	σ_v/σ_ϵ:	**1/2**	1	2	4	8	16
-0.05		**0.0000**	0.0000	0.0001	0.0002	0.0006	0.0009
-0.18		0.0004	0.0018	0.0069	0.0259	0.0826	0.1673
-0.30		0.0038	0.0153	0.0603	0.2296	0.7674	1.7459
-0.43		0.0150	0.0599	0.2374	0.9135	3.1572	7.8451
-0.55		0.0329	0.1313	0.5218	2.0350	7.3811	20.856
-0.68		0.0510	0.2037	0.8119	3.2015	12.097	38.872
-0.80		0.0783	0.3128	1.2485	4.9486	19.083	66.005

$T = 150$, $(\gamma_b, \gamma_f, \lambda) = (0.37, 0.59, 0.015)$, $\rho_1 = 0.9$ and $\sigma_\epsilon = 0.18$.
In bold is the value based on the actual parameter estimates.

The mis-specification parameter

Suppose the structural equation (31) is mis-specified, and the DGP is given by:

$$y_t = \theta_0' Y_t + b' Z_t + u_t$$

When the mis-specified model (31) is asymptotically identified, let θ_p denote the pseudo-true value for a particular GMM estimator (its probability limit). For instance, the pseudo-true value for the efficient 2-step estimator is:

$$\theta_p^{(2)} = \left(\Sigma_{ZY}' V(\theta_p^{(1)})^{-1} \Sigma_{ZY} \right)^{-1} \Sigma_{ZY}' V(\theta_p^{(1)})^{-1} \Sigma_{Zy}$$

which depends on the plim of the 1-step estimator $\theta_p^{(1)}$. The latter is conventionally based on the non-robust weighting matrix Σ_{ZZ}^{-1}, so that $\theta_p^{(1)} = \left(\Sigma_{ZY}' \Sigma_{ZZ}^{-1} \Sigma_{ZY} \right)^{-1} \Sigma_{ZY}' \Sigma_{ZZ}^{-1} \Sigma_{Zy}$. When

we consider local alternatives of the form $b = \bar{b}/\sqrt{T}$, all GMM estimators of θ, under the usual Hansen (1982) regularity conditions, converge to the true value, $\theta_0 = (\Sigma'_{ZY}\Sigma_{ZY})^{-1}\,\Sigma'_{ZY}\Sigma_{Zy}$.

Given θ_p, the mis-specification parameter is:

$$\nu^2 = T\,b'\left[V(\theta_p)^{-1} - V(\theta_p)^{-1}\Sigma_{ZY}\left(\Sigma'_{ZY}\,V(\theta_p)^{-1}\,\Sigma_{ZY}\right)^{-1}\,\Sigma'_{ZY}V(\theta_p)^{-1}\right]\,b. \tag{36}$$

Description of the GG data

The empirical results of this paper are based on the original data set of Galí and Gertler (1999).[39] The data is quarterly, and the sample size is from 1960:Q1 to 1997:Q4. The variable definitions and measurement are given in table 8.

Table 8: GG data description.

Mnemonic	Description	Definition
π_t	Quarterly inflation rate	100 Δ log(GDP deflator).
s_t	Labour share (in deviation from steady state)	$c \times 100 \log \frac{\text{unit labour cost}}{\text{unit price}}$.[a]
gap_t	Output gap	Quadratically detrended real GDP.
spr_t	Long-short interest rate spread	1y bond rate – 3m Fed funds rate.
dw_t	Quarterly wage growth	100 Δ log (unit labour costs).
dc_t	Quarterly commodity price inflation	100 Δ log(commodity prices).

[a] c is a correction factor due to Sbordone (2002): $c = \frac{n(m-1)}{m^2-n}$, where n is the share of labour in the Cobb Douglas production function $Y = A\,K^{1-n}\,L^n$, and m is the average markup of prices over unit costs. GG set $n = 2/3$ and $m = 1.1$, so $c = 0.12$.

References

Abel, A. B. and F. S. Mishkin (1983). An integrated view of tests of rationality, market efficiency and the short-run neutrality of monetary policy. *Journal of Monetary Economics 11*, 3–24.

[39] I thank Prof. Adrian Pagan for kindly sharing this data with me.

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. In *Monetary Policy Rules*. Chicago: University of Chicago Press.

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 H. M. Pesaran (1995). Multivariate rational expectations models: A review and some new results. In Pesaran and Wickens (Eds.), *Handbook of applied econometrics*, Volume Macroeconomics, pp. 139–187. Basil-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 (1985). Staggered wage setting with real wage relativities: Variations on a theme of Taylor. In A. Arbor (Ed.), *Macroeconomic Theory and Stabilization Policy*, pp. 183–199. 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.

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

den Haan, W. J. and A. T. Levin (1997). A practitioner's guide to robust covariance matrix estimation. In *Robust inference*, pp. 299–342. Amsterdam: North-Holland.

Dufour, J.-M. (1997). Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica 65*(6), 1365–1387.

Engle, R. F., D. F. Hendry, and J.-F. Richard (1983). Exogeneity. *Econometrica 51*(2), 277–304.

Ericsson, N. R. and D. F. Hendry (1999). Encompassing and rational expectations: How sequential corroboration can imply refutation. *Empirical Economics 24*, 1–21.

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

Fuhrer, J. C., G. R. Moore, and S. Schuh (1995). Estimating the Linear-Quadratic inventory model: Maximum Likelihood versus Generalized Method of Moments. *Journal of Monetary Economics 35*, 115–158.

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.

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

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

Westview Press.

Hendry, D. F. (1988). The encompassing implications of feedback versus feedforward mechanisms in econometrics. *Oxford Economic Papers 40*, 132–149.

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.

Ma, A. (2002). GMM estimation of the New Keynesian Phillips Curve. *Economics Letters 76*, 411–417.

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

Nelson, C. R. and R. Startz (1990). Some further results on the exact small sample properties of the instrumental variable estimator. *Econometrica 58*(4), 967–976.

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

Pesaran, M. H. (1981). Identification of rational expectations models. *J. Econometrics 16*(3), 375–398.

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

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

Rudd, J. and K. Whelan (2001). New tests of the New-Keynesian Phillips Curve. Discussion Paper, Federal Reserve Board, USA.

Sbordone, A. M. (2002). Prices and unit labor costs: a new test of price stickiness. *Journal*

of Monetary Economics 49, 265–292.

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

Stock, J. and J. Wright (1997). Asymptotics for GMM estimators with Weak Instruments. *NBER Technical Working Paper 198*, 1–1.

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. manuscript, Kennedy School of Government, USA.

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

Wang, J. and E. Zivot (1998). Inference on structural parameters in instrumental variables regression with weak instruments. *Econometrica 66*(6), 1389–1404.

West, K. D. (1997). Another heteroskedasticity- and autocorrelation-consistent covariance matrix estimator. *Journal of Econometrics 76*, 171–191.

Wickens, M. R. (1993). Rational Expectations and Integrated Variables. In P. C. B. Phillips and V. B. Hall (Eds.), *Models, Methods, and Applications to Econometrics, Essays in Honor of Rex Bergstrom*, pp. 317–336. Oxford: Basil Blackwell.