Non-Uniqueness of Deep Parameters and Shocks in Estimated DSGE Models: A Health Warning

Stephen Wright*

September 29, 2012

Abstract

Estimation of dynamic stochastic general equilibrium (DSGE) models using state space methods implies vector autoregressive moving average (VARMA) representations of the observables. Following Lippi and Reichlin's (1994) analysis of nonfundamentalness, this note highlights the potential dangers of non-uniqueness, both of estimates of deep parameters and of structural innovations.

*Department of Economics, Maths & Statistics, Birkbeck College, University of London, Malet Street, London W1E 7HX, UK. s.wright@bbk.ac.uk. I thank Kostas Theodoridis for extremely helpful comments.

1 Introduction

In recent years a burgeoning literature has developed in which the "deep parameters" of dynamic stochastic general equilibrium models are estimated, rather than calibrated. Some models, notably that of Smets and Wouters (2007) have had a significant influence; as has the ready availability, and ease of use, of the estimation and modelling package Dynare, which allows DSGE models to be estimated, and impulse response functions to be derived, almost with a click of a mouse.

This note deals with an aspect of DSGE estimation that appears to have been largely neglected thus far. In virtually all such models, the number of state variables exceeds (often by quite a wide margin) the number of observables. Estimation must therefore always involve a hidden state problem, and is carried out using state-space techniques (frequently exploiting Bayesian priors). A key feature of the solution of such hidden state problems is that the implied process for the observables is always of a finite order vector autoregressive moving average (VARMA) form, with, crucially, non-zero moving average components. As such, these models are potentially subject to the critique of Lippi & Reichlin (1994), that both parameters and driving innovations are non-unique.

Lippi and Reichlin pointed out that, for any given (unique) "fundamental" VARMA representation, there are (usually) multiple alternative "nonfundamental" representations. All such representations must match the time series properties of the observables equally well; but only in the unique fundamental case can the associated innovations at time t be recovered from the t-dated history of the observables. As such, nonfundamental representations have tended to attract little attention from econometricians, since they are non-viable econometric models (see for example the cursory treatment in Hamilton, 1994). While a few macroeconomists have noted the implications of nonfundamentalness (eg, Hansen & Sargent,1993; Fernandez-Villaverde, Rubio-Ramirez, Sargent, and Watson, M, 2007), the Lippi-Reichlin critique had arguably also, until recently, been somewhat sidelined by fact that applied macroeconomists had focussed most of their analysis of the data on pure vector autoregressions, which, if treated as the true data generation process, do not suffer from non-uniqueness.[1]

This note argues that the new approach to DSGE estimation brings the Lippi-Reichlin critique back to the foreground of applied macroeconomics. The parameters of the reduced form VARMAs associated with estimated DSGE models are

[1] Although, as Lippi & Reichlin point out, an estimated VAR may simply be be a convenient approximation for the properties of the true VARMA, thus masking the true non-uniqueness problem.

often quite complicated functions of underlying "deep parameters" (see, for example, Iskrev, 2010). But if the reduced forms are non-unique, this raises the (strong) possibility that the associated deep parameter estimates and driving innovations are also non-unique. Indeed I show that in a benchmark just-identified case this must be so.

There are only two general types of escape route from this problem. First, there may be restrictions on the range of the mapping function from deep parameters, possibly conditional on theory-based restrictions on the parameter space from which the latter are drawn. Second, over-identifying restrictions (which are the norm in DSGE modelling) *may* migitate, or even remove, the problem of non-uniqueness (although I show, in a worked example, that such restrictions may not always do so). But, even if it works, this latter escape route brings with it somewhat uncomfortable implications. The fundamental motivation for estimated DSGE models is that they should match the data better than purely calibrated models. But the analysis of this paper suggests that the problem of non-uniqueness can only be resolved by constraining models such that they fit *less* well.

This note does not attempt a comprehensive analysis of the problem. After setting out some general properties, the bulk of the paper focuses on an extremely simple model, with a single observable, that can be used to demonstrate some of the key factors affecting non-uniqueness, derived from Lettau's (2003) extension of Campbell's (1994) linearised stochastic growth model. It remains to be seen whether the health warnings so derived have general applicability to much more complex estimated DSGE models.

2 The ABCD Structural Model, its VARMA equivalent and Deep Parameters

2.1 The structural model

Writing the underlying linearised DSGE model in its structural and observable form as in Fernandez-Villaverde et al (2007, henceforth FRSW)

$$\begin{aligned} X_{t+1} &= AX_t + Bu_{t+1} \\ Y_{t+1} &= CX_t + Du_{t+1} \end{aligned} \tag{1}$$

X_t is an $r \times 1$ vector of states, u_t is an $s \times 1$ vector of structural innovations, and Y_t is an $n \times 1$ vector of observables. Thus A is $r \times r$, B is $r \times s$, C is $n \times r$, D is $n \times s$. The matrices A, B, C, D are all functions of some vector of deep parameters, $\mathbf{d}$. Given that X_t will usually contain some pre-determined variables, the state dimension r is usually greater than the stochastic dimension s.

It is also standard practice to assume the "square case", in which s is equal to the number of observables, n. Indeed, in estimation it is a requirement that we must have $s \geq n$, since otherwise the covariance matrix of innovations to the observables would be singular. It is less obvious that we must have precise equality of the two dimensions; but a practical consideration is that if we do *not* do so we clearly cannot hope to recover the structural innovations from the data.[2]

The vector of observables may in principle include some subset of non-predetermined variables, like consumption, that solve (by assumption, uniquely) the underlying rational expectations problem, and are a linear function of the states; but it may also in principle include observable elements of, or linear combinations of, the states themselves. In both cases we can therefore assume (with minimal loss of generality) that, underlying (1) we have a levels relationship of the standard measurement equation form[3]

$$Y_t = H'X_t \tag{2}$$

and hence we have $C = H'A$, $D = H'B$. We can also, by appropriate ordering and scaling of the elements of X_t, set $B = \begin{bmatrix} I_s & 0_{s\times(r-s)} \end{bmatrix}'$; hence, partitioning H conformably such that $H' = \begin{bmatrix} H_1 & H_2 \end{bmatrix}$ we have $D = H_1$.

2.2 The Poor Man's Invertibility Condition in the Square Case

Assumption A1 a) $n = s$; b) $|D| = |H_1| \neq 0$; c) All eigenvalues of A lie within the unit circle.

Following FRSW, under Assumption 1 we can write

$$u_{t+1} = D^{-1}\left[Y_{t+1} - CX_t\right]$$

which gives, after substitution into the state equation,

$$X_t = [I - \Gamma L]^{-1} D^{-1} Y_t$$

[2]The requirement that s must be at least as large as n does not necessarily imply that every element of u_t must be regarded as a true structural innovation - some elements may simply be measurement errors (see ref).

[3]While the filtering literature usually includes a measurement error, this can straightforwardly be incorporated into X_t.

where L is the lag operator, and

$$\Gamma = A - BD^{-1}C \tag{3}$$

Hence, as FRSW note, if the eigenvalues of Γ lie within the unit circle (the so-called Poor Man's Invertibility Condition) under A1 the states can be recovered from the history of the observables.

Exploiting results in Baxter, Graham & Wright (2010), we then have

Lemma 1 *$rank(\Gamma) \leq r - n$, with equality if $|A| \neq 0$.*

Proof. See Appendix ■

and using this property we can derive

Proposition 1 ***(The VARMA Representation and the Poor Man's Invertibility Condition)*** *Equation (1) implies that the observables Y_t have a minimal fundamental VARMA(p, q) representation*

$$\Lambda (L) Y_t = \theta (L) \varepsilon_t \tag{4}$$

with autoregressive order $p \leq r - n + 1$, and moving average order $q \leq r - n$, where $\Lambda (L)$ is an $n \times n$ matrix polynomial, but $\theta (L) = \prod_{i=1}^{q} (1 - \theta_i L)$ is a scalar polynomial.

a) If the poor man's invertibility condition holds, then the structural shocks u_t are fundamental for Y_t, i.e.,letting $\gamma_i = eig(\Gamma)$, $i = 1...r - n$,

$$\gamma_i (-1, 1) \ \forall i \Rightarrow \theta_i = \gamma_i \ \forall i, \ \varepsilon_t = H_1 u_t$$

b) Otherwise u_t are innovations to a "structural" nonfundamental representation of the same order of the form

$$\Lambda (L) Y_t = \gamma (L) H_1' u_t \tag{5}$$

where $\gamma (L) = \prod_{i=1}^{q} (1 - \gamma_i L)$. which generates identical time series properties, but u_t and hence X_t cannot be recovered from the history of Y_t. In the fundamental representation (4), for any $\gamma_j \notin (-1, 1)$, $\theta_j = \gamma_j^{-1}$.

Proof. See Appendix ■

While the link between the PMIC and nonfundamentalness is widely known, the precise nature of the link between the dimensions of the structural model and the

VARMA reduced form, and between the eigenvalues of Γ and the MA component, does not appear to be so widely known.[4]

In general there is no theoretical basis for assuming that the PMIC will hold - indeed FRSW present a simple example where theory would predict that it will not. I present another simple case below.

Note that, whether or not the PMIC holds, using Proposition 1, and results in Lippi & Reichlin (1994) the VARMA representation (4) must always have $2^{r-n}-1$ associated "basic" (i.e., of same order) nonfundamental representations, all of which have identical covariance properties. When we work backwards from observable properties and attempt to estimate deep parameters, this multiplicity is crucial.

2.3 Non-uniqueness of deep parameters in the just-identified case.

Let $\mathbf{t} \in \mathbb{T} \subset \mathbb{R}^{\tau}$ be the vector of parameters in the observable VARMA representation (which, it should be recalled, imposes significant restrictions on the general VARMA representation of the same order), and let $\mathbf{d} \in \mathbb{D} \subset \mathbb{R}^{\delta}$ be the deep parameters that determine all matrices in (1). Thus $\mathbf{t} = f(\mathbf{d})$, and hence $\gamma_i = g_i(\mathbf{d})$

Assumption A2: a)*Assume* $\tau = \delta$; b) *that* $\mathbf{d}$ *is always locally identified as in Iskrev (2010); and c) for all* i, *there is some* $d_j \in \mathbb{D}$, *such that* $\gamma_i = g_i(\mathbf{d}_j) \in (-1,1)$, and some $d_k \in \mathbb{D}$, *such that* $\gamma_i = g_i(\mathbf{d}_k) \notin (-1,1)$

We then have straightforwardly:

Corollary 1 *Under A1 and A2, there are* 2^{r-n} *distinct values of* $\mathbf{d}$, *and* 2^{r-n} *distinct sets of shocks* $\{u_t\}_{t=1}^{T}$ *consistent with any given history of the observables* $\{Y_t\}_{t=1}^{T}$

Is the problem of non-uniqueness unavoidable? There appear to be two general ways in which there may be partial, or complete escape routes from the problem.

1. Part c) of Assumption A2 may not hold:

    - For some, or possibly all the γ_i, the mapping from the deep parameters may be restricted to lie solely within $(-1,1)$. If this holds for all i, the PMIC holds. But in general the PMIC does not, directly at least, correspond to any restriction derived from theory;
    - Or it may hold for only some i (which at least reduces the number of deep parameter estimates);

[4] I assume these properties must in fact have been written down somewhere in the structural time series literature; I would be very happy to be told where.

- Or equally, for some, or all i, γ_i may always lie *outside* $(-1, 1)$.

2. In general DSGE models will not be just identified, but will impose over-identifying restrictions. I illustrate below that this *may* result in unique deep parameter estimates. However, the better the DSGE matches the data, the closer it must be to the just-identified case. Thus the pursuit of goodness-of-fit carries with it the risk of (possibly severe) non-uniqueness of deep parameter estimates.

3 Identification of deep parameters in the Campbell (1994)/Lettau (2003) stochastic growth model

I now consider a simple example that illustrates some of the general ideas of of the previous section. This example sacrifices realism in the interest of clarity by assuming that the econometrician wishes to estimate the deep parameters of the simplest possible log-linearised real business cycle model, as in Campbell's (1994), using data for the risk-free rate (exploiting Lettau's (2003) extension of Campbell's framework to include asset prices).

Clearly the mis-matches between this model and the data are so extensive - if they were not, the far more complex models of, eg CEE, Smets & Wouters would presumably never have been developed - that no econometrician would be likely to estimate this system in practice. Nonetheless the model is (just) rich enough to illustrate all the key ideas outlined above for more complex models.

3.1 Model structure

The law of motion of the r =2 states in the Campbell-Lettau model can be written in a manner consistent with (1) as

$$\begin{aligned} X_{t+1} &= \begin{bmatrix} a_{t+1} \\ k_{t+1} \end{bmatrix} = \begin{bmatrix} \phi & 0 \\ \mu & \lambda \end{bmatrix} X_t + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u_{t+1} \\ &= AX_t + Bu_{t+1} \end{aligned} \tag{6}$$

and we can write the process for the $n = 1$ observable, the risk-free rate, as

$$\begin{aligned} y_{t+1} &= \begin{bmatrix} h_a & h_k \end{bmatrix} X_{t+1} = h'X_{t+1} \\ &= h'AX_t + h'Bu_{t+1} = CX_t + Du_{t+1} \end{aligned} \tag{7}$$

where all elements of the matrices are functions of deep parameters. Lettau (2003) provides formulae for h_a and h_k in terms of deep parameters. We then have

$$\Gamma = \begin{bmatrix} -\frac{h_k}{h_a} \\ 1 \end{bmatrix} \begin{bmatrix} \mu & \lambda \end{bmatrix} \tag{8}$$

which has the single eigenvalue, $\gamma = \lambda - \frac{h_k}{h_a}\mu$. Note that the nature of the measurement equation, as well as the state equation, is crucial.

It is then straightforward to show that there will be a fundamental ARMA$(2,1)$ of y_t corresponding to (4) of the form

$$y_t = \left(\frac{1 - \theta L}{(1 - \lambda L)(1 - \phi L)} \right) \varepsilon_t \tag{9}$$

where we have two cases

$$\begin{aligned} \text{Case 1} &: \gamma \in (-1,1) \Rightarrow \theta = \gamma; \varepsilon_t = h_a u_t \\ \text{Case 2} &: \gamma \notin (-1,1) \Rightarrow \theta = \gamma^{-1}; \varepsilon_t = \left(\frac{1 - \gamma L}{1 - \theta L} \right) h_a u_t \end{aligned}$$

thus in Case 2 the observable innovation ε_t will be a lag polynomial function of the true structural innovation; equivalently, $h_a u_t$ will be the innovation to the nonfundamental ARMA$(2,1)$ representation of the same order.

While ϕ, the AR parameter of technology, maps directly through to the ARMA reduced form, γ and λ are both complicated functions of the deep parameters of the model, say

$$\begin{aligned} \lambda &= g_\lambda(\mathbf{d}) \\ \gamma &= g_\gamma(\mathbf{d}) \end{aligned}$$

and hence we have

$$\begin{aligned} \theta &= g_\theta(\mathbf{d}) \\ &= g_\gamma(\mathbf{d}); g_\gamma(\mathbf{d}) \in [-1,1] \\ &= g_\gamma(\mathbf{d})^{-1}; g_\gamma(\mathbf{d}) \notin [-1,1] \end{aligned}$$

3.2 The Just-Identified Case

In Campbell's (1994) original framework the model is calibrated in terms of the steady state return, r, the growth rate g, the labour share α, and the depreciation rate δ : Campbell then analyses the impact of different values of ϕ and σ, the elasticity of intertemporal substitution, on the solution.[5] Even if we (reasonably) treat r and g as known, if we wished to treat the remaining deep parameters $(\alpha, \delta, \phi, \sigma)$ as all unknown, the model would be *under*-identified, since (ignoring constants and variances) a freely estimated ARMA(2,1) has three parameters.[6] At a minimum, therefore, we must fix at least one of the deep parameters. If, for example, we fix δ, the model is just-identified, and we have the mapping

$$\begin{bmatrix} \phi \\ \lambda \\ \gamma \end{bmatrix} = \begin{bmatrix} \phi \\ g_\lambda(\sigma, \alpha) \\ g_\gamma(\sigma, \alpha, \phi) \end{bmatrix}$$

where the parameters of the functions $g_\lambda(.)$ and $g_\gamma(.)$ following Lettau (2003) are (r, g, δ) .These are well-behaved differentiable functions, and we show below that $g_\gamma(.)$ maps to $(0, \infty)$, hence, from Corollary, for freely estimated ARMA parameters $(\phi^*, \lambda^*, \theta^*)$, there will be two distinct implied estimates of (at least) ϕ and σ.[7]

[section to be expanded: should be possible to illustrate geometrically/numerically]

3.3 The Over-Identified Case

Now consider the possibility that we may wish to regard both δ and α as known. We then have over-identification. To illustrate, I shall simplify further by assuming that ϕ is also known and equal to 0.95 (while this figure is of course not known it is often treated as if it is).[8] Thus the two remaining ARMA parameters are both functions of the single remaining deep parameter, which it is helpful to parameterise as $d = 1/\sigma$, the coefficient of relative risk aversion.

Assume further that the likelihood for the ARMA representation can be given the local quadratic approximation

$$l(\lambda, \theta) \approx l(\lambda^*, \theta^*) - \alpha(\lambda - \lambda^*)^2 - \beta(\theta - \theta^*)^2$$

where (λ^*, θ^*) are the unconstrained ML estimates of the two free parameters.

[5] Note that this approach effectively switches off the necessary link, via the Euler equation, that any change in σ should change the steady state return, r.

[6] Counting constants and variances as additional parameters of both structural and ARMA models, the problem is actually worsened.

[7] There is an additional wrinkle, that, given two freely estimated AR parameters, say AR_1^* and AR_2^*, we cannot know, without additional assumptions, which is ϕ, and which is λ.

[8] Allowing ϕ to be estimated would not change matters very much. Campbell (1994) shows that the mapping to λ (η_{kk} in his notation) is invariant to ϕ . Raising (reducing) ϕ makes the $g_\gamma(d)$ function illustrate in Figures 1 to 3 shallower (steeper), but, for $\phi \neq 0$‘, it always crosses unity once.

Figure 1 illustrates two cases where over-identification leads to uniqueness; Figure 2 shows a case where it does not. Both figures[9] assume that $g_\lambda(.)$ and $g_\gamma(.)$ are monotonically increasing in d. We shall see shortly that this is satisfied in this example, but clearly more generally non-monotonicity would complicate matters further.

The top panel of Figure 1 shows that, given the properties of $g_\gamma(d)$, $g_\theta(d)$ must have a peak at $\theta = 1$. In the absence of any other constraints, there would be two values of the deep parameter $d = g_\gamma^{-1}(\theta^*)$, and $d = g_\gamma^{-1}\left(\theta^{*-1}\right)$ associated with the ML estimate of the MA parameter θ^*, with equal associated values of the likelihood.

In the lower panel of Figure 1, consider first the case where the unconstrained ML estimate of the free AR parameter is given by λ_1^*. It follows that the constrained estimate of d must satisfy $\widehat{d} \in \left(g_\lambda^{-1}(\lambda_1^*), g_\gamma^{-1}(\theta^*)\right)$, which will be unique. In effect the sample information on λ forces the constrained estimate of γ to equal $\widehat{\theta}$, the (constrained) fundamental MA parameter. A point estimate of d such that γ was equal to $\widehat{\theta}^{-1}$ would push the associated estimate of λ further away from it ML value, and hence have a lower associated likelihood. In this case, the restricted estimate of the ARMA residual, $\widehat{\varepsilon}_t$, will, up to a scaling factor, provide a direct estimate of the structural error, u_t.

Now consider the case where the unconstrained ML estimate of the free AR parameter is given by λ_2^*. By similar arguments, we must have $\widehat{d} \in \left(g_\gamma^{-1}\left(\theta^{*-1}\right), g_\lambda^{-1}(\lambda_2^*)\right)$. Here the sample information on λ forces γ to equal the constrained *non*fundamental MA parameter, ie $\widehat{d} = \widehat{\theta}^{-1}$, and hence over-identification again brings about uniqueness. However the structural innovation u_t will *not* be a scaling of the ARMA residual, ε_t, but will be as in the second case of (9). It would however be derivable by backward smoothing from full sample information.

Now consider the case where, as in Figure 2, the unconstrained ML estimate of the free AR parameter is given by the intermediate value λ_3^*. In this case the constrained estimates must be non-unique. Figure 2 illustrates 2 deep parameter estimates, $\widehat{d}_1$ and $\widehat{d}_2$, such that

$$\begin{aligned} \left|\lambda_3^* - g_\lambda\left(\widehat{d}_1\right)\right| &= \left|\lambda_3^* - g_\lambda\left(\widehat{d}_2\right)\right| \\ &\text{and} \\ g_\theta\left(\widehat{d}_1\right) &= g_\theta\left(\widehat{d}_2\right) \end{aligned}$$

[9] All figures are appended to the paper.

implying, given the quadratic approximation

$$l\left(\widehat{d_1}\right) = l\left(\widehat{d_2}\right)$$

By inspection of Figure 2, it follows that, given the assumptions on $g(.)$, non-uniqueness will also arise in any sample such that

$$\lambda^* \in \left(g_\lambda\left(g_\gamma^{-1}(\theta^*)\right), g_\lambda\left(g_\gamma^{-1}\left(\theta^{*-1}\right)\right)\right)$$

In all such cases one of the deep parameter estimates will be associated with a (constrained) fundamental ARMA representation, the other with a nonfundamental representation. In such cases the data would simply tell us nothing about which of the two deep parameter estimates to choose.

Can we be sure which of these cases will apply in our example? Figure 3 shows that, without further assumptions, we cannot. Figure 3 uses Campbell's calibration for other parameters[10] to illustrate the counterparts to the functions, $g_\gamma(d), g_\lambda(d)$ (here defined in terms of the only free AR parameter, λ) and $g_\theta(d)$.

The figure shows that for sufficiently low levels of risk aversion the "structural" MA parameter γ is less than unity, but, beyond a critical value ($d = \sigma^{-1} \approx 0.3$) it rises above unity. Beyond this point, therefore, the structural ARMA representation of the risk-free rate must be nonfundamental, and hence in a Campbell/Lettau world the t-dated history of the risk-free rate would not perfectly reveal the technology shock at time t. A higher value of $d = \sigma^{-1}$ also raises the stable eigenvalue of the rational expectations solution, λ, towards unity.

Whether or not estimation of the deep parameter implied by the model would yield unique estimates would, as in Figures 1 and 2, depend on the properties of the data. It is striking that the model's implied fundamental MA parameter, θ, must always lie in a quite narrow range (due largely to the non-monotonicity of the function $g_\theta(d)$); whereas the implied value of λ is distinctly more sensitive to d (especially for low values of d). Thus it seems likely that in this example the properties of λ in the data would largely determine whether the deep parameter estimate was unique. If, for example, the freely estimated value of λ was sufficiently low, this would likely result in a low, and unique, estimate $\widehat{d}$, implying that the technology shock was fundamental. Conversely, a sufficiently high freely estimated value of λ would imply a sufficiently high value of $\widehat{d}$ that it would again be unique (but the technology shock would be nonfundamental). These cases would correspond

[10] $g = 0.005, r = 0.015, \alpha = 2/3, \delta = 0.25$, on quarterly data. The functions shown in Figure 3 are derived in a Maple worksheet, available from the author.

to those illustrated in Figure 1. However, if the data pointed to a freely estimated value of λ in a range around 0.9 , or if the data strongly favoured estimates of θ close to unity, non-uniqueness, as illustrated previously in Figure 2, would be perfectly possible. Thus, to give a quantified example, $d = 0.1 \Rightarrow \theta = 0.99, \lambda = 0.86$, whereas $d = .67 \Rightarrow \theta = 0.99$, $\lambda = 0.95$. If the freely estimated value of λ lay roughly midway between these two values, the data would not be able to discriminate between $\widehat{d} = 0.1$ and $\widehat{d} = 0.67$.

Of course, in this particular example, theory may well be argued to come to our aid. Thus it is commonly (though not universally) assumed that $d = \sigma^{-1} \geq 1$. If we are prepared to make this assumption, either on *a priori* grounds, or on the basis of other evidence, then, by inspection of Figure 3, it is evident that γ must be greater than unity. Thus the bad news, if we made this restriction, would be that the Poor Man's Invertibility Condition does not hold; but the good news would be that our estimate of d would be unique. But it is hard to know how much consolation to draw from this: it is not obvious that more complex models would be so obliging.

This example has focussed on the properties of a particular observable, the risk-free rate. Campbell's original analysis focussed on the properties of consumption. An intriguing property of this example is that, for the parts of the parameter space in which the PMIC does not hold for the risk-free rate, it *does* hold for consumption. The reverse also holds. Thus if both consumption and the risk-free rate were observable in the Campbell-Lettau framework, either innovations to consumption would predict the risk-free rate, or vice versa.

Of course in practice we would expect both (true) consumption (certainly) and the risk-free rate (less certainly) to be measured with error. However, even if true consumption could not be directly observed, if the risk-free rate is observed without error, and $\gamma > 1$, the *properties* of the true consumption process can be inferred from the properties of the nonfundamental counterpart to (9). Furthermore, in any finite sample, we can always derive a "backward" smoothed estimate $u_{t|T}$, that may enable us to say quite a lot about the properties of the true structural shock, with a greater precision, the bigger is the difference between T and t.

4 Conclusions

This note seeks to bring the attention of those engaged in estimating dynamic stochastic general models to the potential pitfalls of non-uniqueness, both of deep parameter estimates, and of estimates of structural innovations that arise from non-fundamentalness, as first pointed out by Lippi & Reichlin (1994). The analysis has been purely illustrative, in a highly simplified framework. But it seems unlikely that non-uniqueness would be *less* of a problem in richer models, since, as models get larger, the multiplicity of nonfundamental solutions increases very rapidly.

It is worth stressing that nonfundamentalness, as such, is not a problem, at least as far as deep parameter estimates are concerned. If we can pin down a unique, albeit nonfundamental representation for the observable, as would arguably be the case for our example, then the deep parameter estimate is unique; the problem then is confined to that of uncovering structural shocks using backward smoothing.

Finally it might be tempting to conclude that the problem of non-uniqness can at least be contained, if not necessarily eliminated if, for high r (the number of states) we also have high n (the number of observables and shocks, given the assumption that $s = n$) since, as the analysis of this note shows, if $r - n$ is low, the multiplicity of parameter estimates is at least reduced. However, this seems too convenient. In most estimated DSGEs, it is already debatable whether s, the relatively large number of structural shocks - which must be assumed to be equal to n, the number of observables, can really be rationalised as true structural shocks. “Solving” the problem of non-uniqueness by introducing yet more notionally structural shocks would appear an unfortunate approach to anyone with even a nodding acquaintance with Occam’s Razor.

Appendix

A Proof of Lemma

Using (2) we have

$$\Gamma = A - BD^{-1}C = A - B\left[H'B\right]^{-1}H'A = \left(I - B\left[H'B\right]^{-1}H'\right)A$$

but pre-multiplying by H' we have

$$H'\Gamma = H'\left(I - B\left[H'B\right]^{-1}H'\right)A = 0$$

hence since H' is $n \times r$, Γ can have at most $r - n$ non-zero eigenvalues.■

B Proof of Proposition

Writing

$$\begin{aligned} X_t &= \left[I - \Gamma L\right]^{-1} BD^{-1}Y_t \\ &= \left[I - \Gamma L\right]^{-1} B\left(H'B\right)^{-1}Y_t \end{aligned}$$

given, that, as stated in main text, we have

$$B = \begin{bmatrix} I_n \\ 0 \end{bmatrix}$$

then

$$\begin{aligned} \begin{bmatrix} u_t \\ 0 \end{bmatrix} &= \left[I - AL\right]X_t \\ &\text{hence} \\ u_t &= B'\left[I - AL\right]X_t \\ &= B'\left[I - AL\right]\left[I - \Gamma L\right]^{-1} B\left(H'B\right)^{-1}Y_t \end{aligned}$$

and hence

$$\det\left(I - \Gamma L\right)u_t = B'\left[I - AL\right]\operatorname{adj}\left[I - \Gamma L\right]B\left(H'B\right)^{-1}Y_t$$

if we premultiply by $H'B = H_2$ where $H' = \left[\begin{array}{cc} H_1 & H_2 \end{array} \right]$ and $H_1 \equiv D$ is, by assumption $n \times n$ and invertible, we can write this as

$$\Lambda(L) Y_t = \gamma(L) H_1 u_t$$

where $\gamma(L) = \det(I - \Gamma L) = \Pi_{i=1}^{r-n}(1 - \gamma_i L)$ and $\Lambda(L) = I + \Lambda_1 L + \Lambda_2 L + ... + \Lambda_p L^p$. Thus $\gamma(L)$ is a common *scalar* MA polynomial of order $q \leq r - n$ and $\Lambda(L)$ is, also as in **??**, of order $p \leq r - n + 1$, since all elements of adj$(I - \Gamma L)$ are of order q. Note that $\Lambda(0) = I_n$, since

$$\Lambda(L) = H_1 B' [I - AL] \operatorname{adj} [I - \Gamma L] B H_1$$

and, since

$$adj (I - \Gamma L) = \det(I - \Gamma L)(I - \Gamma L)^{-1} = \Pi_{i=1}^{r-n}(1 - \gamma_i L)\left(I + \Gamma L + \Gamma^2 L^2 + ...\right)$$

$\Lambda(0)$ is given by

$$H_1 B' B H_1^{-1} = I_n$$

We then have two cases:

a) If the PMIC holds $\gamma_i \in (-1.1)$ $\forall i$, and hence $\theta_i = \gamma_i$ $\forall i$, and hence $\varepsilon_t = H_1 u_t$

b) If the PMIC does not hold, $\gamma_j \notin (-1, 1)$, for some j, $\theta_j = \gamma_j^{-1}$, but for any $\gamma_i \in (-1, 1)$, $\theta_j = \gamma_j$. In all such cases u_t will be the innovations to a nonfundamental representation.■

References

Campbell, John Y, (1994), "Inspecting the mechanism: an analytical approach to the stochastic growth model", Journal of Monetary Economics 33, pp.463-506.

Fernandez-Villaverde, J, Rubio-Ramirez, F, Sargent, T and Watson, M (2007) "ABCs (and Ds) of Understanding VARs", American Economic Review, pp 1021-1026

Hamilton J D (1994) *Time Series Analysis* 1994 Princeton University Press

Hansen, L P and Sargent, T J (1993), *Recursive Linear Models of Dynamic Economies*, unpublished

Harvey, Andrew (1989) "Forecasting, Structural Time Series Models and the Kalman Filter" Cambridge University Press

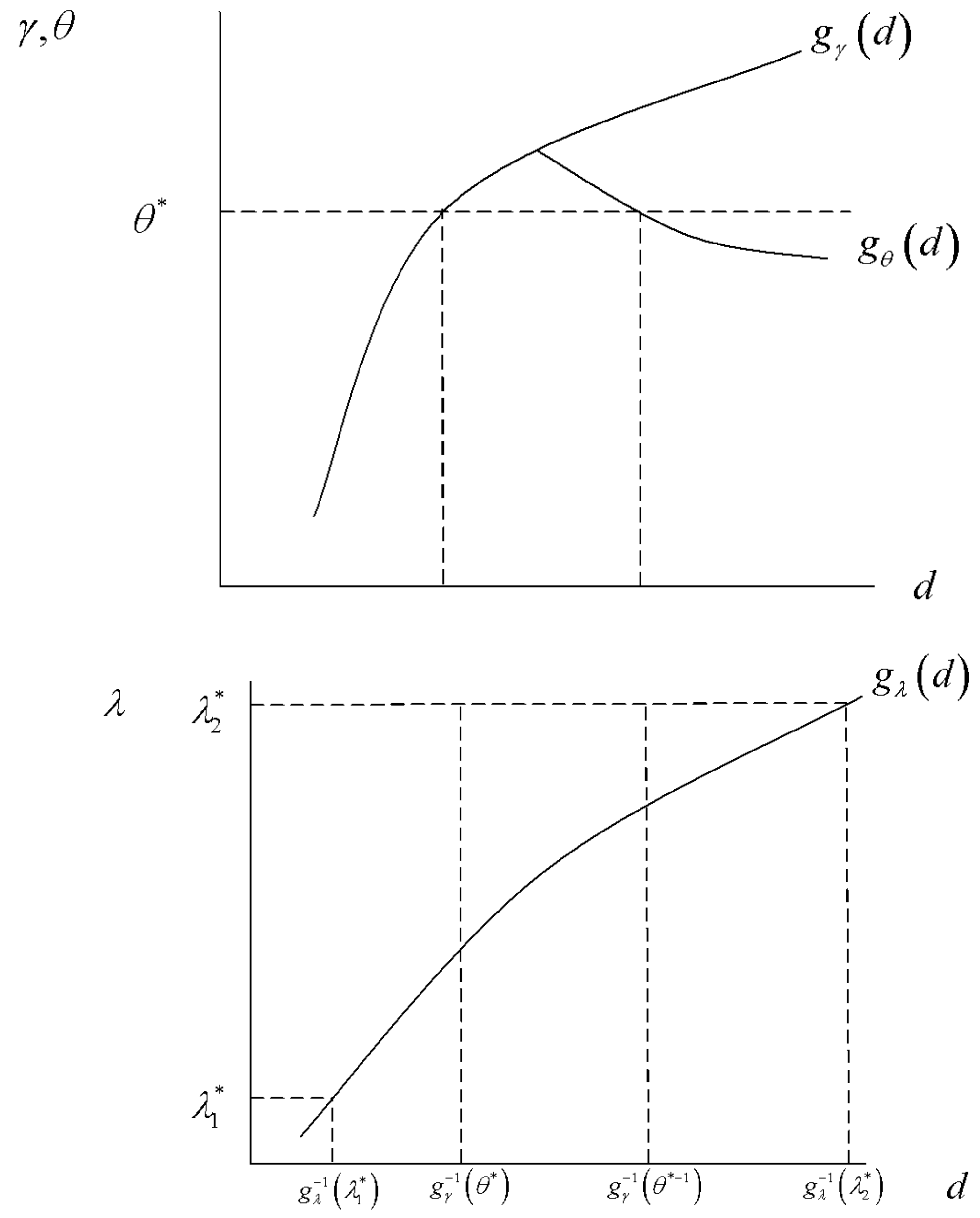
Iskrev, Nikolay (2010). "Local identification in DSGE models," Journal of Monetary Economics, Elsevier, vol. 57(2), pages 189-202

Lettau, Martin (2003), “Inspecting the Mechanism: Closed-Form Solutions for Asset Prices in Real Business Cycle Models", *Economic Journal,* 113 550–575

Lippi, Marco and Reichlin, Lucrezia (1994) "VAR analysis, nonfundamental representations, Blaschke matrices", Journal of Econometrics, 63 pp 307-325

Smets, F and Wouters, R (2007) “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach”, *American Economic Review,* 97 , pp 586-606

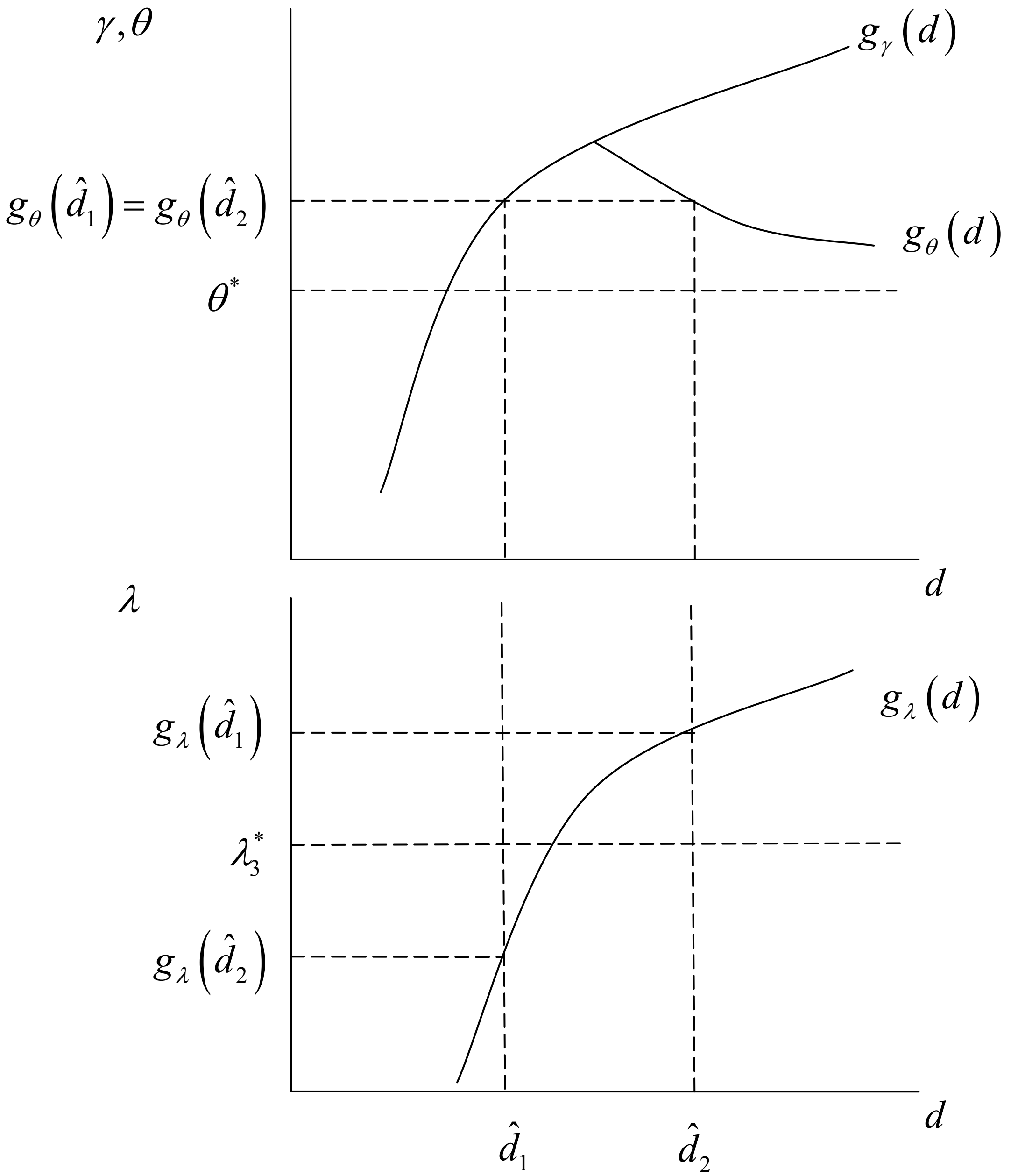
Figure 1.
Uniqueness of Deep Parameter Estimates in an Over-Identified ARMA(2,1)



Note to Figure 1:

Given freely estimated MA parameter θ^* and (free) AR parameter λ_1^* the constrained estimate of the deep parameter $\hat{d}$ will lie between $g_\lambda^{-1}\left(\lambda_1^*\right)$ and $g_\gamma^{-1}\left(\theta^*\right)$; if the freely estimated AR parameter equals λ_2^*, $\hat{d}$ will lie between $g_\gamma^{-1}\left(\theta^{*-1}\right)$ and $g_\lambda^{-1}\left(\lambda_2^*\right)$. In both cases the deep parameter estimate will be unique.

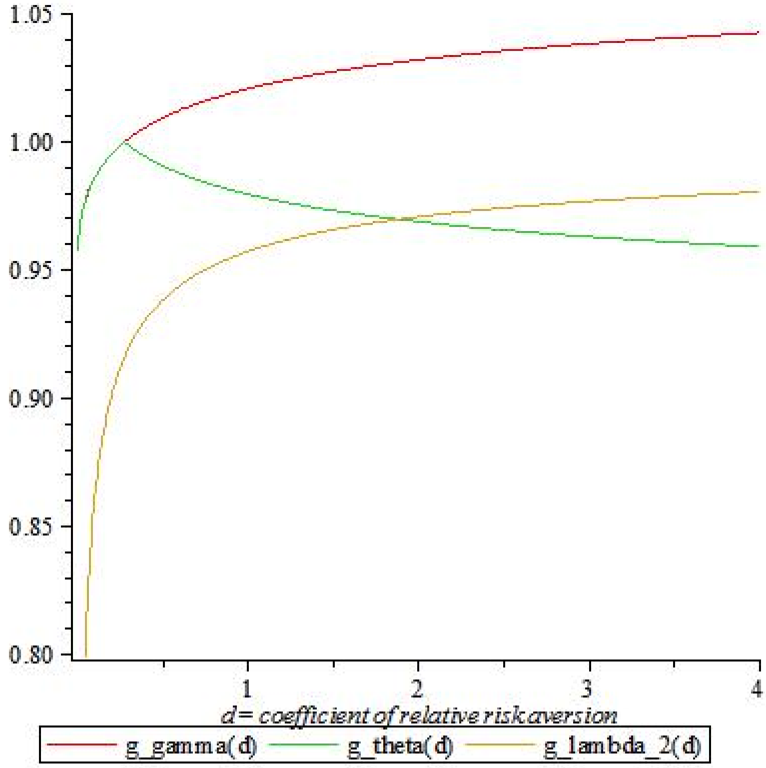
Figure 2.
Non-Uniqueness of Deep Parameter Estimates in an Over-Identified ARMA(2,1)



Note to Figure 2:

At deep parameter point estimates, $\hat{d}_1$ and $\hat{d}_2$, the implied values of θ are identical, and the two implied values of λ are equidistant from the unconstrained value λ_3^*, hence to a quadratic approximation the likelihood is equal at both estimates: $l(\hat{d}_1) \approx l(\hat{d}_2)$

Figure 3:
Implied ARMA(2,1) Parameters for the risk-free rate in the Campbell (1994)/Lettau (2003) log-linear stochastic growth model.



Note to Figure 3

Figure 3 illustrates the implied mapping, in the log-linearised stochastic growth model, from a single deep parameter, *d*= the coefficient of relative risk aversion, to the free ARMA(2,1) parameters of the risk-free rate, using Campbell's (1994) calibration for other deep parameters. The first AR parameter, φ is given by the AR(1) parameter of the technology shock, set equal to 0.95.