

Invertible and Non-Invertible Information Sets in Dynamic Stochastic General Equilibrium

Brad Baxter*, Liam Graham† and Stephen Wright‡

March 2, 2010

Abstract

Most DSGE modelling derives rational expectations solutions by assuming that all state variables relevant to optimising decisions are directly observable, or that they are "...an invertible function of observables" (Mehra and Prescott, 1990). In this paper, we derive a number of results useful for the analysis of DSGE models with restricted information sets. We make a distinction between instantaneous and asymptotic invertibility, and show that the latter may require significantly less information than the former. We also show that non-invertibility of the information set can have significant implications for the time-series properties of economies.

JEL classifications: D80; E32.

Keywords: imperfect information; invertibility; fundamental versus non-fundamental time series representations; Kalman filter; dynamic stochastic general equilibrium

*School of Economics, Maths & Statistics Birkbeck College, University of London, Malet Street, London W1E 7HX, UK. b.baxter@bbk.ac.uk

†Department of Economics, University College London, Gower Street, London WC1E 6BT, UK. Liam.Graham@ucl.ac.uk

‡Corresponding author: School of Economics, Maths & Statistics Birkbeck College, University of London, Malet Street, London W1E 7HX, UK. s.wright@bbk.ac.uk

1 Introduction

Underlying most dynamic stochastic general equilibrium models is the strong assumption that all state variables relevant to optimising decisions are directly observable. This can be rationalized, as in Mehra and Prescott (1990) by the assumption that the states are “...an invertible function of observables”.

In this paper we consider a general linear DSGE model in which optimising agents use the Kalman filter to estimate the state variables from their information set. Both states and observables may in principle be endogenous to the optimising decisions made in response to this information. We distinguish between three kinds of information set, examine their properties and show that the nature of the information set can have important implications, both for the model’s rational expectations solution and for the time series properties of the economy

1. **Instantaneously invertible information sets.** The states can be recovered from only t - dated information, the Kalman filter is redundant and the full information solution is replicated.

2. **Asymptotically invertible information sets.** In the limit as the history of observables becomes infinite, the Kalman Filter converges to a solution which replicates full information. We show that the conditions for asymptotic invertibility can be satisfied with fewer observables than needed for instantaneous invertibility and put important additional restrictions on the *nature* of the measurement process, and its interaction with the dynamics of the underlying states.

3. **Non-invertible information sets.** The Kalman filter can never replicate the full information solution. We show that the impact of limited information is transitory, but can be highly persistent, and that the endogenous nature of the filtering problem introduces new (but unobservable) sources of dynamics in response to structural shocks. The only observable dynamics will however be those of the estimated, as opposed to the actual states. We show that, with non-invertible information sets, the estimated states will follow the same vector autoregressive process as would the true states in a notional full information economy, but with a different covariance pattern of shocks. In this notional economy, pre-determined variables like capital are subject to “pseudo-shocks” that are logically impossible under full information.

We also examine the implications of information for the time series properties of the economy. Invertibility of an information set is closely related to invertibility or fundamentalness, in time series analysis. The solution to a Kalman filter problem *always* results in a fundamental time series representation of the observables, for which the innovations can be recovered from the history of the observables. But

invertibility of the information set implies that the structural innovations that drive the true states are also fundamental in the time series sense. In contrast, under non-invertible information sets the structural innovations cannot be recovered from the history of the observables. Indeed, in the special case of a non-invertible information set in which the number of observables equals the number of structural innovations, we show that the latter are also (up to a scaling factor) innovations to a *nonfundamental* time series representation of the observables. We also show that in general the endogenous nature of the filtering problem leads to nonbasic (Lippi and Reichlin, 1994) nonfundamental representations of the observables.

To complement the paper we provide a Matlab toolkit which allows the easy application of our techniques to a wide class of linear rational expectations models.

Our analysis builds on past analysis of macroeconomic models with limited information (Pearlman et al, 1986; Pearlman, 1992; Svensson and Woodford, 2003, 2004). It also exploits known results in control theory (e.g., Whittle, 1983, Söderström, 1994) and in econometric inference in time series representations (Hansen and Sargent, 2005; Lippi and Reichlin, 1994; Fernandez-Villaverde et al, 2007), but which have not, to our knowledge, been applied to the problem we address here.

Limited information sets within macroeconomic models may arise, as in, e.g. Svensson and Woodford (2003) and Pearlman (1992), where a policymaker sets policy variables with incomplete information on the underlying state variables in the economy, or, as in Bomfim (2004), Keen (2004), Collard and Dellas (2006) where representative consumers are assumed to face informational restrictions. Woodford (2003), Lorenzoni (2010), Nimark (2007), Graham and Wright (2010) and Poračkarm and Young (2008) address the problem of heterogeneous agents facing a symmetric filtering problem of inferring aggregates which requires them to form estimates of a “hierarchy of average expectations“ (Townsend, 1983). Graham and Wright (2010) show that when such models include endogenous states a filtering problem of the form discussed in the present paper is central to the solution. We discuss this particular case further in Section 7.2.

The remainder of the paper is organized as follows. Section 2 summarises the key properties of the general filtering problem within a general DSGE model. In Section 3 we give conditions for asymptotic invertibility and in Section 4 we discuss non-invertible information sets. Section 5 discusses implications for time series properties. In Section 6 we show how our techniques can be applied to an analytical example based on the benchmark stochastic growth model. Section 7 discusses applications and extensions of our results. Appendices provide proofs and algebraic derivations.

2 The signal extraction problem

2.1 A general system representation

A linearised structural dynamic stochastic general equilibrium model can be written as in McCallum (1998) as:

$$A_{yy}E_t y_{t+1} = B_{yy}y_t + B_{yk}k_t + B_{yz}z_t \quad (1)$$

$$k_{t+1} = B_{ky}y_t + B_{kk}k_t + B_{kz}z_t \quad (2)$$

$$z_{t+1} = B_{zz}z_t + \zeta_{t+1} \quad (3)$$

In the first block of equations y_t is a $q \times 1$ vector of non-predetermined variables. The matrix A_{yy} may not be invertible. The equations in this first block may be first-order conditions (both static and dynamic), identities or production relations. The second block describes the evolution of an $r_k \times 1$ vector of predetermined variables, k_t , while the third describes the evolution of an $r_z \times 1$ vector of exogenous stochastic processes, z_t , that can be represented by a first order vector autoregression, with ζ_t an $r_z \times 1$ vector of iid innovations with covariance matrix $S_{zz} = E(\zeta_t \zeta_t')$ which we assume is full rank.¹ Using results such as those in Hamilton (1994, pp 678-9) a process as in (3) can also represent a wide range of stochastic processes such as Markov Chains by an appropriate specification of the process for the innovations.

We assume that agents form expectations based on an information set

$$I_t = \{\{i_{t-j}, j \geq 0\}; \Xi; \Phi_0\} \quad (4)$$

where Ξ contains the (time-invariant) structure and parameters of equations (1) to (4), Φ_0 represents initial estimates of the states and their covariance matrix², and i_t is an $n \times 1$ vector of observed variables

$$i_t = C_{ik}k_t + C_{iz}z_t + C_{iy}y_t + C_{iw}w_t \quad (5)$$

w_t is an $r_w \times 1$ vector of measurement errors, with $0 \leq r_w \leq n$.³

For generality we can in principle allow these to be serially correlated by repre-

¹Higher order VARMA representations of exogenous variables may in principle be captured by including lags of z_t and current or lagged values of u_t in k_t , and allowing z_{t+1} to depend on k_t . With this small amendment there is no loss of generality in assuming that S_{zz} is full rank.

²One value for Φ_0 would be the unconditional properties of the states implicit in Ξ . Under conditions given in Proposition 2, initial estimates will have no impact on the equilibrium.

³Measurement errors may be of lower dimension than the measured variables themselves if, for example, some linear combination of k_t , z_t and y_t is measured without error, or if measurement errors in different variables are systematically related.

senting them as a vector autoregression:

$$w_{t+1} = B_{ww}w_t + \omega_{t+1} \quad (6)$$

where ω_t has the full rank covariance matrix $S_{\omega\omega} = E[\omega_t\omega_t']$. We assume that the eigenvalues of B_{ww} have real parts less than or equal to unity. The two vectors of innovations ω_t and ζ_t may in principle be contemporaneously correlated, with $E(\zeta_t\omega_t') = S_{\zeta\omega}$, but are assumed uncorrelated at all other leads and lags.

The nature of the information set (which will be model-specific) should ideally be explicitly related to the underlying structure of the economy.⁴ It should also be borne in mind that, as we show in Appendix A, the structure of the model in (1) to (3) needs to be both informationally consistent and informationally feasible i.e. any linear combinations of the states appearing in the forward looking equations of (1) must be observable.

We provide a Matlab toolkit which takes as input a system in the form specified in equations (1) to (6), and implements all the transformations and solution methods that follow⁵.

2.2 The filtering problem

For compactness of notation we incorporate predetermined and exogenous variables, k_t and z_t , together with measurement error, w_t , into a vector of state variables ξ_t of dimension $r = r_k + r_z + r_w$. We partition y_t into two sub-vectors, x_t with variables given by static relationships and c_t with observable dynamic choice variables that satisfy expectational difference equations such as consumption or policy variables. Note that none of our results depend on this partition.⁶

In Appendix A we show that we can then use (2) to (5) to derive the following compact representation of the state evolution and measurement equations:

$$\xi_{t+1} = F_\xi\xi_t + F_c c_t + v_{t+1} \quad (7)$$

$$i_t = H'_\xi\xi_t + H_c c_t \quad (8)$$

⁴For example, Graham and Wright (2010) argue that, in an incomplete markets setting, information should be assumed to be "market-consistent" - ie agents only obtain information from the markets they trade in. Porapakkarm and Young (2008) is another example of this approach.

⁵Available from the authors' websites.

⁶This is discussed in more detail in Appendix A.

where

$$\xi_t = \begin{bmatrix} k_t \\ z_t \\ w_t \end{bmatrix}; v_t = \begin{bmatrix} 0 \\ \zeta_t \\ \omega_t \end{bmatrix}$$

$$Q = E(v_t v_t') = \begin{bmatrix} 0_{r_k \times r_k} & 0_{r_k \times s} \\ 0_{s \times r_k} & S \end{bmatrix}; S = \begin{bmatrix} S_{\zeta\zeta} & S_{\zeta\omega} \\ S'_{\zeta\omega} & S_{\omega\omega} \end{bmatrix}$$

Definitions of the matrices in (7) and (8) are given in Appendix A.

We assume that optimising agents solve the signal extraction problem using the Kalman filter. The system in (7) and (8) differs from that underlying standard derivations of the Kalman filter used in econometrics (as in, e.g., Harvey, 1981; 1989; Hamilton, 1994; Hansen and Sargent, 2005) where the system is usually written as

$$\xi_{t+1} = F_\xi \xi_t + v_{t+1} \tag{7a}$$

$$i_t = H'_\xi \xi_t + w_t \tag{8a}$$

where ξ_t are the state variables. Comparing this with the system in (7) and (8) reveals a number of differences that are of significance for our results.

First, there are two forms of endogeneity in the system in (7) and (8) since both equations depend on the dynamic choice variables c_t . An important distinction between the two forms of endogeneity is that the states have only lagged dependence on c_t (the most obvious example being the impact of current consumption decisions on future capital) while the measured variables may have contemporaneous dependence (for example via intratemporal optimality conditions). Both forms of endogeneity have important implications for the nature of the filtering problem, and its consequences for the economy in which it takes place.⁷

Second, in econometric applications of the Kalman filter, where the ξ_t are exogenous state processes, it is typically assumed that these are either stationary or at worst may have unit roots. Thus the eigenvalues of the matrix F_ξ are assumed to be not greater than unity in absolute value. In contrast, in the endogenous Kalman filter problem generated by a typical dynamic stochastic general equilibrium model in (7), F_ξ will have at least one explosive eigenvalue, due to the dynamics of capital under dynamic efficiency.⁸ We shall show that this feature interacts in interesting ways

⁷Endogeneity of this form makes the filtering problem in the context of a DSGE much closer to those analysed in control theory, as in eg, Whittle, 1983; Anderson and Moore, 1979; Soderstrom, 1994.

⁸We give an example of a standard model (based on the linearisation in Campbell, 1994) that has this property in Section 6.

with the endogeneity of the Kalman filter, and indeed *requires* some endogeneity, if any latent explosive roots are to be stabilised.

Third, Q , the covariance matrix of the innovations of the redefined states, defined after (8), is of rank $s = r_z + r_w \leq r$, with the inequality holding in strict form when there are pre-determined variables (if $r_k > 0$).

Fourth (and more trivially) the measurement errors, w_t , have been absorbed into the redefined states, ξ_t . This allows us to accommodate, in principle, both serial correlation of measurement errors and contemporaneous correlation with the structural innovations, u_t ; it also allows some simplifications of the key formulae.

While we have derived the filtering problem from a standard dynamic stochastic general equilibrium model, in which the different elements of the state variables have a clear interpretation (and imply a number of restrictions on the structure of the problem) in the remainder of the paper the only features of the system in (7) and (8) that are crucial to our results are the overall state dimension, given by r , the stochastic dimension given by $s \leq r$, and the number of measured variables, $n \leq r$, along with the endogeneity of both states and measured variables to the dynamic choice variables, c_t .

2.3 Full information solution

The special case of full information provides a crucial analytical building block for the more general solution under other information sets.

Definition 1 (*Full information*) *Full information implies the state variables are known, $i_t = \xi_t$.*

Full information is a special case of the system (7) and (8) with $n = r$, $H_\xi = I_r$, $H_c = 0$. The Kalman filter is therefore redundant.

The solution for the dynamic choice variables under full information that satisfies the expectational difference equations in (1) can then be expressed in the form

$$c_t^* = \eta' \xi_t^* \tag{9}$$

where for any variable x_t , x_t^* denotes its value under full information. The matrix η , depends in general on all structural and preference parameters of the model, and can be computed using standard techniques (e.g. Blanchard and Kahn, 1980 ; McCallum, 1998).⁹ For the rest of the paper we assume it to be unique and treat

⁹Note that all elements in the i th row of η are zero for $i > r_k + r_z$ (measurement errors have no impact on c_t under full information).

it as parametric. Given (9), the full information states follow a first order vector autoregressive process in reduced form:

$$\xi_{t+1}^* = G\xi_t^* + v_{t+1} \quad (10)$$

where

$$G = F_\xi + F_c\eta' \quad (11)$$

The behaviour of the dynamic choice variables, c_t , under full information turns out to be crucial for the stability of the states under any information set. As noted in the previous section, in a model with endogenous capital F_ξ will usually have at least one explosive eigenvalue. This latent explosive property can only be controlled by the behaviour of the dynamic choice variables. Under full information this stabilisation follows directly from the standard rational expectations solution. Under standard conditions the matrices η and F_c always satisfy the following conditions:

Assumption 1. *All the eigenvalues of the matrix $G = F_\xi + F_c\eta'$ have real parts less than or equal to unity*

Assumption 2. *Let $G = V\Lambda V^{-1}$ where Λ is a diagonal matrix of eigenvalues and V the corresponding matrix of eigenvectors. For any strictly unit eigenvalue in Λ the corresponding row of F_c is zero.*

Assumption 1 rules out explosive rational expectations solutions;¹⁰ Assumption 2 states that, to the extent that any innovations have permanent effects, these are innovations to strictly exogenous processes (e.g., there may be a unit root component in technology).¹¹ These features of the solution under full information turn out to be equally crucial for the stability of the solution even when the information set is non-invertible, and thus does not replicate full information.

2.4 Instantaneous invertibility

The full information solution is replicated under the following conditions:

Definition 2 (*Instantaneous invertibility*). *An information set is instantaneously invertible if $n = r$, the number of observables is equal to the state dimension, and*

¹⁰Note that there is no direct link between the explosive eigenvalues in the rational expectations solution of (1) to (3) (which are set to zero in solving the model and deriving η under full information), and those of F_ξ , except to the extent that, by inspection of (11), the stability of the rational expectations solution requires that the explosive eigenvalues of F_ξ be “switched off” by the behaviour of consumption.

¹¹Assumption 2 follows naturally from the underlying structural model in (1) to (3), since the dependence of state variables on c_t is only via k_{t+1} , hence all elements of the i th row of F_c are zero for $i > r - s$.

H_ξ in the measurement equation (8) is invertible, and hence full information can be replicated using only t - dated information.

In this case the state variables can be obtained by inverting the measurement equation (8) to give

$$\xi_t = (H'_\xi)^{-1} (i_t - H_c c_t) \quad (12)$$

hence again the Kalman filter is redundant, as is all lagged information on the observables $\{i_{t-j}\}$ for $j > 0$.

2.5 The Kalman filter

For the remainder of the paper we shall assume that the information set is not instantaneously invertible. We therefore need to apply the Kalman filter.

Given the linear structure of the model we follow standard practice¹² and conjecture that optimal choices will be “certainty-equivalent” i.e.

$$c_t = \eta' \widehat{\xi}_t \quad (13)$$

where $\widehat{\xi}_t = E\xi_t|I_t$ is the optimal estimate of the current state vector given the available information set I_t and η is identical to the matrix for the full information case in (9). We show in the discussion of Corollary 2 that this conjecture is verified.

We first define two key matrices that characterise the properties of the state estimates and state forecasts.

$$M_t = E \left[\left(\xi_t - \widehat{\xi}_t \right) \left(\xi_t - \widehat{\xi}_t \right)' \right] \quad (14)$$

is the covariance matrix of the filtering errors of the state estimates, and

$$P_{t+1} = E \left[\left(\xi_{t+1} - E_t \xi_{t+1} \right) \left(\xi_{t+1} - E_t \xi_{t+1} \right)' \right] \quad (15)$$

is the covariance matrix of the one-step ahead state forecast errors.¹³

The following proposition shows that despite the endogeneity of both states and measured variables to the dynamic choice variables, the solution to the filtering problem can be related to standard results:

¹²See, for example, Pearlman (1992); Svensson and Woodford (2004), who in turn are simply applying standard results in control theory as in, eg, Whittle (1983), Soderstrom (1994)

¹³For compactness of notation we write the period t estimate of the states as $\widehat{\xi}_t$; whereas the standard Kalman filter literature commonly uses $\widehat{\xi}_{t|t}$. For the forecast at time t of the states at period $t+1$ we write $E_t \xi_{t+1}$ ($= E_t \widehat{\xi}_{t+1}$) instead of the standard $\widehat{\xi}_{t+1|t}$. P_{t+1} is commonly denoted $P_{t+1|t}$, and using the same notation $M_t = P_{t|t}$, but we separate the two for clarity.

Proposition 1 (*The parallel problem*). *In the solution to the “endogenous” Kalman filter problem given by (7) and (8), the mean squared error matrices P_{t+1} and M_t are identical to those derived from a parallel exogenous Kalman filter problem*

$$\begin{aligned}\tilde{\xi}_{t+1} &= F_\xi \tilde{\xi}_t + v_{t+1} \\ \tilde{i}_t &= H'_\xi \tilde{\xi}_t\end{aligned}$$

(i.e., setting $F_c = H_c = 0$ in (7) and (8)) so are given by the standard Kalman filter recursion

$$P_{t+1} = F_\xi M_t F'_\xi + Q \quad (16)$$

$$M_t = \left[I_r - \tilde{\beta}_t H'_\xi \right] P_t \quad (17)$$

$$\tilde{\beta}_t = P_t H_\xi \left[H'_\xi P_t H_\xi \right]^{-1} \quad (18)$$

In the solution to the actual problem, the estimated states follow the process

$$\hat{\xi}_{t+1} = G \hat{\xi}_t + \beta_t \varepsilon_{t+1} \quad (19)$$

where G is as defined in (11), ε_t is the vector of innovation to the observables,

$$\varepsilon_{t+1} \equiv i_{t+1} - E_t i_{t+1} \quad (20)$$

and

$$\beta_t = \tilde{\beta}_t \left[I_n + H_c \eta' \tilde{\beta}_t \right]^{-1} \quad (21)$$

Proof. See Appendix B. ■

The filtering problem set out in equations (7) and (8) displays two forms of endogeneity: the dependence of the states on the lagged dynamic control variables, via F_c in (7) and the dependence of the measured variables on the contemporaneous dynamic control variables, via H_c in (8). Proposition 1 states that the solution to this problem can be derived from the solution to a parallel filtering problem for a notional state process $\tilde{\xi}$ and a notional set of measured variables \tilde{i}_t for which both effects are absent, so that the standard exogenous Kalman filter formulae can be applied.¹⁴

We can derive the Kalman filter from the properties of the parallel, rather than the true process, because, while the dynamic choice variables determine future states in the true problem via the matrix F_c , this does *not* impact on one-step ahead

¹⁴Our formulae for P_{t+1} and $\tilde{\beta}_t$ are more compact than the more common formulation, given our absorption of measurement error into the states, but can be easily shown to be identical.

uncertainty (since the marginal impact of today’s choices on tomorrow’s states is known today even if current states are unknown). As a result the expression for P_{t+1} only allows for the direct impact of uncertainty about today’s states transmitting to uncertainty about tomorrow’s states, via the matrix F_ξ . Since the matrix F_c does not affect the solution to the filtering problem, it can be solved under the assumption that $F_c = 0$.

The Kalman gain¹⁵ matrix β_t for the true problem is not however the same as its counterpart $\tilde{\beta}_t$ in the parallel problem because the signal conveyed by innovations to the measured variables also affects the dynamic choice variables c_t . But this has no impact on the mean squared error matrices M_t and P_{t+1} , hence these can be derived under the assumption that $H_c = 0$.

Proposition 1 shows that, conditional upon the solution for β_t in (21), the estimated states $\hat{\xi}_t$ follow the process given by (19) with the same non-explosive autoregressive matrix G as for the true states under full information, in (10). In the parallel problem, in contrast, the notional state process $\tilde{\xi}_t$ has autoregressive matrix F_ξ , which, as noted above may have explosive eigenvalues. We shall see below that this rather unusual feature of the state process in the parallel problem has interesting implications for the nature of information processing.

In Appendix B we show that the solution to the filtering problem given by Proposition 1 is implied by standard results in control theory (see, for example, Söderström, 1994, Ch. 6); it is also equivalent to that in Svensson and Woodford (2003) and in Pearlman et al (1986). However the specification in terms of the parallel problem has, as far as we know, thus far gone unremarked; and it is this feature that provides the basis for many of the results that follow.

2.6 The steady state Kalman filter

Equations (16) to (18) are a set of recursive matrix equations, for which it is natural to look for a stable steady state. The re-statement of the endogenous Kalman filter problem in Proposition 1 in terms of the parallel exogenous problem means that this becomes an application of existing results.

¹⁵We use the definition of the Kalman gain as in, eg, Harvey (1981), in which it can be interpreted as a matrix of regression coefficients updating current state estimates in response to forecast errors in predicting measured variables (Soderstrom (1994) refers to this as the “filter gain”) . The term is also frequently applied (as in for example, Hamilton, 1994) to a matrix, often denoted K , that updates *forecasts* of the states in response to the same forecast errors (Soderstrom (*op cit*) refers to this definition as the “predictor gain”). In the parallel exogenous problem $\tilde{K} = F_\xi \tilde{\beta}_t$ in our notation, however in the actual endogenous problem $K = G\beta_t$, since it would incorporate the endogenous response of dynamic choice variables both in β_t but also in the autoregressive representation in (19).

Proposition 2 (*steady-state Kalman filter*). *If the parallel problem in Proposition 1 is stabilisable and detectable in the sense of Anderson and Moore (1979), then for any initial positive definite matrix P_0 , a unique stable steady state of the endogenous Kalman filter problem exists, with matrices β , P and M that satisfy the steady state of equations (16) to (18).*

Proof. See Appendix C. ■

The twin conditions of stabilisability and detectability can both be related to the nature of the shock processes driving the underlying DSGE model. If the innovations to the state process in (7) are expressed in the form

$$v_t = F_u u_t \tag{22}$$

where $F_u = \begin{bmatrix} 0_{(r-s) \times s} \\ I_s \end{bmatrix}; u_t = \begin{bmatrix} \zeta_t \\ \omega_t \end{bmatrix}; E(u_t u_t') = S$

(where the first block of zeros reflects the presence of pre-determined variables) then the two conditions can be written as

$$\text{stabilisability: } |\Lambda_i(F_\xi + F_u L_1')| < 1 \quad \forall i$$

$$\text{detectability: } |\Lambda_i(F_\xi + L_2 H_\xi')| < 1 \quad \forall i$$

where the two conditions are satisfied for some matrices L_1 and L_2 of dimensions $r \times s$ and $r \times n$ respectively, and where $\Lambda_i(A)$ denotes the i th eigenvalue of a matrix A . Note that these conditions apply to the parallel problem, and hence are entirely unaffected by endogeneity in the true filtering problem.

The first condition is trivial if there are no pre-determined variables, and hence u_t the underlying innovations in (22) are of dimension $s = r$, since in that case F_u is a full rank $r \times r$ matrix. Where there are pre-determined variables ($s < r$) it is not so straightforward. In this case F_u will usually contain a row of zeros in exactly the row corresponding to an explosive eigenvalue in F_ξ , so that the condition for stabilisability can only be met if the relevant row of F_ξ contains off-diagonal elements. A simple example might be that capital must depend not only on lagged consumption, but also on lags of stochastic exogenous state variables (for example technology).

The second condition requires that there must be some observable indicator, however poor, of any state variables with associated explosive or unit eigenvalues.

Note that, as is standard in all applications of the Kalman filter, the recursion in equations (16) to (18) does not depend on the data. It is common practice to solve the model using steady-state values of β and P . It should be borne in mind, however,

that this methodology makes the implicit assumption that there is a sufficiently long history of i_t in the information set for the Kalman filter to have converged.¹⁶

3 Asymptotic invertibility

We have derived the Kalman filter for a general signal extraction problem where we have assumed only that the information set is not instantaneously invertible as in Definition 2 (if it were, as we saw in Section 2.4, the Kalman filter would be redundant). In general the information set will be non-invertible: i.e. state estimates will differ from the true states. However, under certain conditions we now show that the Kalman filter will converge to a steady state which replicates full information.

Definition 2 showed that instantaneous invertibility requires that n , the number of measured variables, equal r , the number of states. In most standard exogenous Kalman filter problems, in which the stochastic dimension, s , equals the state dimension, r , this is the only way that full information can be replicated. However, the endogenous nature of the states in in the context of a DSGE model implies that this condition is sufficient, but not necessary.

The dynamic endogeneity of some of the states (the k_t) in the DSGE framework means that, because they are pre-determined, the stochastic dimension s , is less than r , the state dimension. Under conditions which we shall state below, there may as a result be cases in which the Kalman filter can become arbitrarily close to replicating full information as the history of the information set increases over time. This property can be defined formally in three ways that are all logically equivalent.

Definition 3 (*asymptotic invertibility*) *An information set I_t is asymptotically invertible if*

$$M \equiv \lim_{t \rightarrow \infty} M_t(I_t) = 0 \iff P \equiv \lim_{t \rightarrow \infty} P_t(I_t) = Q \iff \lim_{t \rightarrow \infty} [\xi_t^* - \xi_t(I_t)] = 0$$

where, from Proposition 2, M and P satisfy the steady state of the recursion in (16) and (17).

The definition explicitly notes the dependence of M_t , P_t and ξ_t on the entire history of observables, I_t . If the information set satisfies this definition, M_t , the

¹⁶While the recursion does not depend on the data, it *does* depend on the passage of time, reflecting the fact that away from steady state (ie, conditional upon some initial estimate of the states) the covariance matrix of innovations to the measured variables is time-varying: $var_t(\varepsilon_t) = H'P_tH$, and strictly greater than at steady state, with implications for the optimal value of the Kalman gain matrix, β_t .

mean squared error matrix of state filtering errors converges to a steady state value of zero; and P_t , the covariance matrix of one-period-ahead state forecast errors, tends to its irreducible minimum of Q . Given the nature of the recursion in for M_t and P_t in (16) and (17) asymptotic invertibility depends potentially on the entire structure of the model and its interaction with the measurement process.

The first two conditions taken in isolation require only that the estimated states must converge on the true states, whereas the third condition appears stronger, since it states that the true states in turn must converge on their full information values (initial errors in estimating the states must have strictly transitory impact). However we show below that the three features are in fact logically identical.¹⁷

This implies significant restrictions on the nature of the information set, and its relation to the underlying structural model, which we summarise in the following proposition:

Proposition 3 (*conditions for asymptotic invertibility*). *Assume that the Kalman filter of Proposition 1 satisfies the conditions for convergence given by Proposition 2. Necessary and sufficient conditions for an information set I_t to be asymptotically invertible are*

1. $n = s$
2. $|H'_\xi F_u| \neq 0$
3. $|\Lambda_i \left(\left(I - \tilde{\beta}(Q) H'_\xi \right) F_\xi \right)| < 1 \quad \forall i$

where n is the number of observed variables; $s = \text{rank}(S) = \text{rank}(Q)$ is the "stochastic dimension" of the state variables; H_ξ and F_u are as given in equations (8) and (22); $\Lambda_i(A)$ are the eigenvalues of a matrix A ; and $\tilde{\beta}(Q) = F_u (H'_\xi F_u)^{-1}$ satisfies (18) setting $P = Q$.

Proof. See Appendix E. ■

To illustrate, assume that all three conditions in Proposition 3 hold, and that there is a sufficiently long history of the observables that state estimates in period t have converged on their true values. Since all the conditions relate to features of the parallel problem as defined in Proposition 1 we can ignore both forms of endogeneity and set $F_c = H_c = 0$.¹⁸ Manipulation of equations (7), (8), (20) and

¹⁷See Corollary 3 below. Note however that the states will usually converge to their full information values more slowly than P_t and M_t because past errors will have a persistent effect on the true states (see discussion below in Section 4).

¹⁸This assumption is made purely to simplify the algebra; the same qualitative properties hold with both forms of endogeneity.

(22) then implies that in period $t+1$, the innovations ε_{t+1} in the observables (defined in (20)) relate to the underlying structural innovations, u_{t+1} by

$$\varepsilon_{t+1}|_{\xi_t=\widehat{\xi}_t} = H'_\xi \left(\xi_{t+1} - E_t \xi_{t+1} |_{\xi_t=\widehat{\xi}_t} \right) = H'_\xi v_{t+1} = H'_\xi F_u u_{t+1} \quad (23)$$

So, if the first two conditions hold, such that $H'_\xi F_u$ is both square (given $n = s$) and invertible, then if the Kalman filter reveals the true states in period t , the structural innovations can be derived from the innovations to the observables by inverting (23), and hence the states will also be known in period $t + 1$, and so on indefinitely.

However, this simply tells us that if the first two conditions in Proposition 3 are satisfied, then $P = Q$ is a fixed point of the Kalman filter. The third condition tells us whether this fixed point is stable, and shows that the *nature* of the measurement process, and its interaction with the dynamics of the underlying states, is crucial. Recall that the assumption underpinning equation (23) was that the true states were known in period t . But we shall show below¹⁹ that if the third condition in Proposition 3 does not hold, then any errors in state estimates in period t will be locally explosive in the neighbourhood of full information, thus rendering the assumption invalid, since state estimates can never converge on full information.²⁰

The conditions set by Proposition 3 state the minimal assumptions necessary for the assumption of full information to be at least asymptotically valid given a sufficiently long history of observables in the information set. All three conditions can be related to interesting economic features of DSGE models.

1. Asymptotic invertibility is only of interest in the case that s , the stochastic dimension of an economic system, is lower than r , the dimension of the states. The benchmark stochastic growth model, for example, is driven by a single stochastic process for technology. The first condition in Proposition 3 implies that in this economy (subject to the other two conditions also holding) full information can be replicated, at least asymptotically, with only a single observable variable. We illustrate this property in Section 6.
2. A long - recognized inference problem, dating back at least to Muth (1961), and re-visited more recently by Bomfim (2004), arises when technology is subject to shocks with different persistence properties: e.g. if there are both transitory

¹⁹See discussion after Corollary 3.

²⁰In Appendix E we show that the three conditions in Proposition 3 have a precise mathematical counterpart to an invertibility condition for an econometrician stated in Fernandez-Villaverde et al (2007).

and persistent shocks. Since such shocks have an identical initial impact on technology, they will also typically affect all observable variables identically on impact. In such cases, even if the first condition in Proposition 3 is satisfied, so that there are as many observable variables as shocks, the second condition will not be satisfied (i.e., $|H'_\xi F_u| = 0$). Thus inference problems of this type will never disappear, however long the history of the observables.

3. Even if both the first two conditions are satisfied, the third condition, relating to the nature of the measurement process, can also be crucial. In Section 6 we give an example, again in the context of the stochastic growth model, in which the information set satisfies conditions 1 and 2, but does not satisfy condition 3, and therefore is not asymptotically invertible.

4 Non-invertible information sets

We now address the implications of an information set being non-invertible. We first show that even in this case the observable dynamics of the economy can be represented by those of a notional full information economy, but with a different covariance pattern of innovations; we then examine the implied true (but, in real time, intrinsically unobservable) dynamics of the economy, which are more complex.

4.1 An isomorphic representation

Corollary 1 *Assuming convergence of the Kalman filter, the behaviour of the estimated states $\widehat{\xi}_t$ is isomorphic to the behaviour of the true states, ξ_t under full information, if Q , the covariance matrix of underlying structural innovations in (7), is replaced by the matrix $Q + F_\xi M F'_\xi - M$. In this representation there will in general be pseudo-shocks to states that are in reality pre-determined.²¹*

Proof. See Appendix C ■

This property follows directly from the representation of the state estimates in (19), which shows that the $\widehat{\xi}_t$ have an autoregressive representation which is identical that of the full information states, ξ_t^* in (10) except in terms of its innovations. The innovations to the estimated states, which are a linear combination of the innovations to the observables, are vector white noise conditional upon the t - dated information set, and, if the Kalman filter has converged to its steady state, these innovations have a time-invariant distribution. And, given the certainty equivalent nature of

²¹Note that Corollary 1 subsumes both instantaneous and asymptotic invertibility as special cases since both imply $M = 0$.

the optimal choices of the dynamic choice variables in (13) the estimated states are sufficient for a time series representation of the dynamic choice variables, just as the true states are under full information.²²

However, despite its identical autoregressive representation, the notional full information economy described by Corollary 1 will have different dynamic properties, due to the nature of its innovations. Most notably, as stated in Corollary 1, there will in general be pseudo-shocks to predetermined variables that are logically impossible under full information. Thus analysis of an economy with a non-invertible information set may require the derivation of impulse responses to, for example, "capital shocks" that are never considered in the standard analysis of models where agents actually have full information.²³

Given certainty equivalence, any differences in the dynamics of the states in the isomorphic representation in Corollary 1 in turn determine the impact of non-invertible information sets on the time series properties of dynamic choice variables. The additional terms in the innovation covariance matrix of the isomorphic representation do not always sum to a positive definite matrix. Hence non-invertibility of an information set can in principle result in a process for dynamic choice variables with higher, or lower, variance than under full information.²⁴

4.2 True state dynamics and the implications of non-invertibility

While there is, as stated in Corollary 1, an autoregressive representation of the state estimates in terms of observable innovations, this representation does not describe the dynamics of the *true* states, except of course in the special case of invertibility. For the general non-invertible case the true (but unobservable) dynamics of the economy are richer once expressed in terms of the true structural innovations. Analysis of these dynamics provides further important insights.

Define the vector of state filtering errors as

$$f_t \equiv \xi_t - \hat{\xi}_t \tag{24}$$

We show in Appendix D that, conditional upon convergence of the Kalman filter, the

²²This reduced form representation of the estimated states has a precise econometric parallel in the "Innovations Representation" of Hansen and Sargent (2005, p 191).

²³In reality these are of course not true shocks, but an amalgam of true shocks and revisions to past estimates of the states: but Corollary 1 implies that the responses of the dynamic choice variables, and hence of the states, to these pseudo-shocks will be identical to the full information response, if such shocks were actually possible.

²⁴A point also made by Pearlman et al (1986) and Pearlman (1992), and relevant to the results of Bomfim (2004).

joint process for f_t and the true states ξ_t can be expressed in the vector autoregressive form

$$\begin{bmatrix} \xi_{t+1} \\ f_{t+1} \end{bmatrix} = \begin{bmatrix} G & -F_c \eta' \\ 0 & (I - \tilde{\beta} H'_\xi) F_\xi \end{bmatrix} \begin{bmatrix} \xi_t \\ f_t \end{bmatrix} + \begin{bmatrix} I \\ I - \tilde{\beta} H'_\xi \end{bmatrix} v_{t+1} \quad (25)$$

The top block is entirely independent of filtering parameters, and transparently reduces to the full information process (10) for invertible cases in which filtering error disappears. In general, however, with dynamic endogeneity, ($F_c \neq 0$) filtering error “contaminates” state dynamics via the off-diagonal element of the autoregressive matrix for the joint process for ξ_t and f_t . In contrast the process for the vector of state filtering errors f_t is block recursive. Furthermore, consistent with Proposition 1, f_t follows an identical process to the state filtering error in the parallel exogenous problem (i.e., it does not depend on F_c or H_c) and hence is also invariant to the properties of the c_t , the dynamic choice variables.

Proposition 2 has an important corollary that is crucial to the time series properties summarised in (25):

Corollary 2 *In the autoregressive representation (25), the autoregressive matrix $(I - \tilde{\beta} H'_\xi) F_\xi$ of the filtering errors, $f_t \equiv \xi_t - \hat{\xi}_t$, has at most $r - n$ non-zero eigenvalues, all of which have real parts strictly less than unity in absolute value.*

Proof. See Appendix C. ■

Thus convergence of the Kalman filter to a unique steady state automatically implies that filtering errors are stationary (and vice versa).

Corollary 2 also sheds light on the third condition for asymptotic invertibility in Proposition 3. If the first two conditions are satisfied there is a fixed point of the Kalman filter that replicates full information. But, using Corollary 2 the third condition in Proposition 3 can be interpreted as a condition on the eigenvalues of the autoregressive matrix of filtering errors in the neighbourhood of the solution that replicates full information. If these eigenvalues are not stable then, even if full information is a fixed point of the Kalman filter, it will not be stable. State estimates in period t will be locally explosive in the neighbourhood of full information so estimates of the states can never converge on their true values. Assuming the conditions in Proposition 2 hold, there will still be a stable steady state Kalman filter, in the neighbourhood of which, from Corollary 2, filtering errors will be stationary, but it will *not* be one that replicates full information.

The joint process for ξ_t and f_t in (25) provides a complete description of the true process for the dynamic choice variables c_t , since, using (13) we can write $c_t = \eta' \hat{\xi}_t \equiv \eta' (\xi_t - f_t)$: thus with a non-invertible information set the process for c_t

differs from the process under full information both because of the direct effect of filtering error on the estimated states and because the true states differ from their full information values.²⁵ The more persistent is the filtering error process (the closer are the non-zero eigenvalues of $(I - \tilde{\beta}H'_\xi)F_\xi$ to unity), the more prolonged will be the additional dynamics introduced by the filtering problem.

Note that, while filtering errors will contaminate the true state dynamics, they will not do so in expectation, since $E_t f_t = 0$ by definition. Given this, at any point in time the *expected* process for the states has an identical form to that of the true states under full information, hence the undetermined coefficients problem to which η is the solution is identical to that under full information. This feature verifies the assumption of certainty equivalence used in specifying the optimal values of the dynamic choice variables, c_t , in (13).²⁶

The representation in (25), together with Corollary 2, also implies a number of further features of non-invertible information sets that are direct corollaries of Propositions 1 and 2, given Assumptions 1 and 2.

Corollary 3 *Non-invertibility of an information set has no permanent effects, even when there are permanent structural shocks (i.e., if B_{zz} has one or more unit eigenvalues).*

Proof. See Appendix D ■

This Corollary implies that impulse responses under full information and non-invertible information must converge. Thus while short-run dynamics may be significantly affected by informational problems, longer-run responses (especially those to permanent shocks) derived from models that assume full information will approximate those from models with non-invertible information sets. This has potentially important implications for econometric analysis: it means, for example, that identifying assumptions for vector autoregressive models based on long-run restrictions predicated on theoretical models assuming full information are equally applicable under imperfect information.

Corollary 4 *The filtering errors f_t satisfy $H'_\xi f_t = 0$.*

²⁵Note that only filtering errors in the underlying states k_t and z_t have any direct impact on c_t since, as noted previously, η has zeros in its i th row for $i > r_k + r_z$.

²⁶Certainty equivalence is of course a standard result that arises from the fact that we first linearise the model (including Euler equations) and then solve the filtering problem. To the extent that state uncertainty introduces new sources of variance in dynamic choice variables (an issue we discuss in Section 4.3) incorporation of state uncertainty into the optimisation problem before linearisation would presumably result in effects analogous to those in the precautionary consumption literature.

Proof. See Appendix D ■

Linear dependence between the elements of the vector of filtering errors arises from the efficient use of the structural knowledge of the economy that underpins the Kalman filter. To see the intuition for this result, note that, if we take t -dated expectations of the measurement equation (8), using (9) this implies

$$\begin{aligned} i_t &= H_\xi \xi_t + H_c \eta' \widehat{\xi}_t = (H_\xi + H_c \eta') \widehat{\xi}_t \\ &\Rightarrow H_\xi' \xi_t = H_\xi' \widehat{\xi}_t \Rightarrow H_\xi f_t = 0 \end{aligned} \quad (26)$$

thus agents know that filtering errors for any given state variable must be precisely offset by some combination of other filtering errors. By implication neither the innovation matrix of the vector of filtering errors, f_t , nor its autoregressive matrix can be of full rank, and thus the n -dimensional vector f_t can always be expressed in terms of a sub-vector of dimension $r - n$.

Corollary 5 *Let $\beta = \begin{bmatrix} \beta_k & \beta_z & \beta_w \end{bmatrix}'$. If F_ξ has explosive eigenvalues, $\beta_k = 0$ can never be a convergent solution of the Kalman filter problem.*

Proof. See Appendix D ■

This result follows directly from the stationarity of the filtering error process. When F_ξ has explosive eigenvalues associated with the evolution of the pre-determined variables, k_t , the autoregressive matrix of the filtering error in (25), $(I - \widetilde{\beta} H_\xi') F_\xi$, could not have stable eigenvalues with $\widetilde{\beta}_k = 0$. Since, from Corollary 2, stability of the filtering errors is directly equivalent to convergence of the Kalman filter, this in turn implies that in this case $\beta_k \neq 0$ can never be a solution of the filtering problem. This has interesting implications for the nature of the optimal response to information, as the quality of that information deteriorates.

In standard exogenous Kalman filter problems, in which $F_c = 0$ and F_ξ usually has at worst borderline unit eigenvalues, the lower the quality of the information, the smaller is the optimal response to that information. The same feature applies in our problem when F_ξ has stable or unit eigenvalues.²⁷ In contrast, if F_ξ has explosive eigenvalues, so does the state process in the parallel problem of Proposition 1, so for β_k sufficiently close to zero, the filtering error process would itself be explosive, contradicting Corollary 2. In this case, as $S_{\omega\omega}$ tends to infinity, β_k tends to a fixed, non-zero matrix. Thus however poor the information, it is always optimal to respond to it.²⁸

²⁷This feature is noted in Svensson and Woodford (2003, p711) where it appears to be presented as a general result.

²⁸Since this feature derives from the parallel problem it clearly also applies to any exogenous

4.3 A caveat on impulse responses

The joint process for ξ_t and f_t in (25) discussed in the previous section shows that non-invertible information introduces more complicated dynamics than under full information. However, impulse responses derived from the full reduced form representation in (25) would not be observable in real time. The only observable impulse responses would be those to the isomorphic representation of Corollary 1, in which, as already discussed, the innovation covariance properties of this notional economy may be very different from those of the true structural shocks.²⁹

5 Time series representations of the observables

It is a standard result (see Hamilton, 1994, p 391) that, conditional upon convergence of the Kalman filter, the innovations ε_t to the observable variables i_t must also be the innovations to an invertible, or fundamental, time series representation of i_t . The key characteristic of fundamental innovations is that they can be recovered from, and are white noise conditional upon, the history of i_t . The link with the Kalman filter is evident from the fact that the same applies for the innovations that condition upon state estimates, $\hat{\xi}_t$.

Another standard result in time series analysis (Hamilton, 1994, pp 67-68; Lippi and Reichlin, 1994) is that for any fundamental autoregressive moving average representation of an observable vector process such as i_t there is a finite set of alternative representations of the same ARMA order, but each with a different set of nonfundamental innovations. These nonfundamental representations have the same autocovariance properties as the fundamental representation, but the associated innovations are *not* recoverable from the history of i_t .

A further feature of time series representations noted by Lippi and Reichlin (1994) is that for any fundamental representation there is a potentially infinite set of nonfundamental representations of arbitrarily higher ARMA order, which they term non-basic. Lippi and Reichlin conclude however that such non-basic representations

Kalman filtering problem where the autoregressive representation is explosive. There appears to be little discussion of this feature in the existing Kalman filter literature, presumably because explosive representations are rather unusual. For example, Anderson and Moore (1979) dismiss even the borderline unit root case as of limited interest. The closest parallel that we can find is in a worked example in Whittle (1983) p 126, where, for a control problem with a potentially explosive state process, optimal responses are never zero; but interestingly in the context of optimal control rather than optimal filtering. However, given the duality of the filtering and control problems the result is conceptually very close.

²⁹There is a clear counterpart here with the econometric invertibility problem in Fernandez-Villaverde et al (2007).

are “not likely to occur in models based on economic theory” (Lippi and Reichlin 1994, p. 315).³⁰

These characteristics of the time series representation of i_t are invariant to the nature of the information set or of the underlying structural model. Where the nature of the information set and the structural model *does* make a difference is in the nature of the link between the alternative time series representations of i_t and the underlying structural innovations, u_t to the true state process. The following proposition summarises the nature of this link.

Proposition 4 (*time series representations and structural innovations*) For any $n \times 1$ vector of observables, i_t that satisfies the measurement equation (8), and given convergence of the Kalman filter to its steady state:

a) For any information set I_t there exists a fundamental vector autoregressive moving average VARMA(n, p, q) representation of i_t with autoregressive order $p \leq r - n + 1$ and moving average order $q \leq r - n$.

b) If I_t is invertible, the structural innovations u_t are, up to a scaling factor, the innovations to this fundamental representation.

c) If I_t is non-invertible and $n = s = \text{rank}(H'_\xi F_u)$ the u_t are, up to a scaling factor, innovations to a non-fundamental representation of i_t

d) If I_t is non-invertible and $n < s$, the innovations to nonfundamental representations can in general only be recovered from an information set consisting of both the history and future of the structural innovations, u_t .

e) If I_t is non-invertible and there is dynamic endogeneity ($F_c \neq 0$), the true process for i_t implies a nonbasic nonfundamental VARMA representation with reciprocal autoregressive roots given by the eigenvalues of G and the eigenvalues of $(I - \tilde{\beta}H'_\xi)F_\xi$.

Proof. See Appendix F ■

³⁰The intuition for non-basic representations can be demonstrated by noting that any univariate white noise process can in principle be written as a nonfundamental ARMA(1,1) in terms of another white noise process with lower variance. For example, if ω_t is univariate white noise, for any arbitrary $\psi \in (-1, 1)$ we can write

$$\omega_t = \left(\frac{1 - \psi^{-1}L}{1 - \psi L} \right) \zeta_t$$

with $\sigma_\zeta^2 = \psi^2 \sigma_\omega^2$. The fundamental representation has MA parameter equal to ψ , and hence the MA and AR components cancel, which indeed they must, since ω_t is univariate white noise. Thus the the history of ω_t can be consistent with any value of ψ and hence any value of σ_ζ^2 . Furthermore, for any arbitrary choice of $\psi \neq 0$ the nonfundamental innovation ζ_t can in turn be given a nonfundamental representation, and so on ad infinitum. Hence in principle there is an infinite set of nonfundamental representation of ω_t , of arbitrary order, but they are all nonbasic.

Table 1 summaries the properties of the information set and the resulting time series representation of i_t .

Part a of the proposition states that the fundamental time series representation of the observables has autoregressive and moving average orders reducing in n , the dimension of the observables. In the special case of instantaneous invertibility ($r = n$) Table 1 shows that both the states and the observables have a first order VAR representation, the reciprocal autoregressive roots of which are the eigenvalues of G , defined in (11).

Part b covers the cases of both instantaneous and asymptotic invertibility. In the first case, we can recover the true states, ξ_t from (12), substitute for c_t , and re-write the law of motion for ξ_t (7) as a vector autoregression for i_t . In the second case of asymptotic invertibility (as in Proposition 3), since the u_t can be recovered from the ε_t , then we can re-write the fundamental VARMA representation by substituting for ε_t in terms of u_t using (23). In both cases, since u_t can be recovered from the history of the observables, i_t , invertibility of the information set and fundamentalness of the u_t are two sides of the same coin.

If the information set is non-invertible, the vector of structural innovations u_t *cannot* be recovered from the history of the observables. Part c of highlights an interesting special case: when the number of structural innovations equals the number of linearly independent observables, but the information set is not asymptotically invertible (because the third condition in Proposition 3 is not satisfied), a VARMA representation of i_t with innovations given by some scaling of u_t does exist, but these innovations must be *nonfundamental*. The nonfundamentalness of u_t in this case is similarly just another way of saying that, despite the fact that both sets of innovations have the same dimension, the u_t cannot be recovered from the history of the observables. As a consequence, there is no static, invertible transformation of the form $\varepsilon_t = Au_t$ for some matrix of scalars, A , as there is in the limiting case of asymptotic invertibility. Instead we have $\varepsilon_t = A(L)u_t$, where $A(L)$ is a matrix polynomial in non-negative powers of the lag operator that maps a vector white noise process to another vector white noise process of the same dimension. In this case the mapping is provided by the true state process represented in (25), but it is non-invertible using the history of the observables. The inversion of $A(L)$ would result in terms in the forward operator, defined such that, $Fx_t = L^{-1}x_t$, hence to recover u_t requires information on the future, as well as the history of ε_t , and hence of the observables.

This linkage between innovations is less straightforward in the more general non-invertible case, summarised in part d. As in the special case in part c, the

true process provides a mapping between the structural and observable innovations: $\varepsilon_t = A(L)u_t$, but $A(L)$ is no longer square since the structural innovations are of higher dimension than the observable innovations ($s > n$). Any nonfundamental VARMA representation also has a set of innovations η_t of the same dimension as ε_t such that $\varepsilon_t = B(L)\eta_t$, where $B(L)$ is square, and also contains only non-negative powers of the lag operator. Thus $\eta_t = B(L)^{-1}A(L)u_t$, where the inversion of $B(L)$ again involves term in $L^{-1} = F$, and hence η_t depends in general on both the history and the future of the structural innovations. Only in the special case in part c, where $n = s$, do we have $B(L) = A(L)$, so that these additional dynamics cancel out.

Part e reveals an aspect of the time series properties of the observables that is specific to the endogenous Kalman Filter. As discussed in Section 4, the true state process in (25) contains additional dynamics due to the contamination of the state process by filtering error due to dynamic endogeneity (when $F_c \neq 0$). These additional dynamics are not observable: so the fundamental reduced form VARMA representation is of the same order as the reduced form of the system under full information (since it can be written in terms of the isomorphic state representation of Corollary 1). However, there is a nonfundamental VARMA representation that contains these additional dynamics: as a result this representation is not only non-fundamental, but also nonbasic in Lippi and Reichlin's (1994) terminology. Thus, in contrast, to Lippi and Reichlin's conclusion that nonbasic nonfundamental representations are unlikely to occur in models based on economic theory, part e shows that, on the contrary, they are an intrinsic feature of models that combine endogenous states with non-invertible information sets. When the states are exogenous, however (ie, if $F_c = 0$), this feature disappears. The last two rows of Table 1 illustrate the contrast between the cases of exogenous and endogenous states.

The nonbasic nature of these representations is not just of theoretical interest. If nonfundamental representations are at least basic, this implies that, while the structural innovations themselves cannot be derived from the history of the observables, it is at least possible for an econometrician to identify from the data some of the *characteristics* of these unobservable structural innovations, from the characteristics of the observable fundamental representation: for example, some of their covariance properties, and some qualitative features of impulse responses.³¹ The non-basic characteristic of VARMA representations that arise from non-invertible

³¹For example, Robertson & Wright (2010) show that observable properties of basic nonfundamental representations can be used to put bounds on the variances of the observables, conditioning on the true states. Lippi & Reichlin (1994) analyse the properties of impulse responses in basic nonfundamental representations, which, as Proposition 4c shows, nest one representation in terms of the true structural innovations, but, from Proposition 4e, this is only the case if the states are exogenous.

information sets with endogenous states rules out even this possibility.

6 Example: information in the stochastic growth model

In this section we present an analytical example that shows how our techniques can be applied to the benchmark stochastic growth model. We consider three aggregate information sets corresponding to the three cases of instantaneous invertibility, asymptotic invertibility and non-invertibility.

Following Campbell (1994) the aggregate state evolution equations are

$$k_{t+1} = \lambda_1 k_t + \lambda_2 a_t + (1 - \lambda_1 - \lambda_2) c_t \quad (27)$$

$$a_{t+1} = \phi a_t + u_{t+1} \quad (28)$$

where capital k_t is pre-determined, a_t is aggregate technology and c_t is consumption. The first equation comes from combining the linearised forms of the aggregate capital evolution equation and the aggregate resource constraint³².

Implicitly underlying the aggregates is a heterogeneous agent model in which each agent faces a budget constraint of the form (27) but in idiosyncratic variables. We do not need to give details of this model (although we discuss it further in section 6.4) since we are only concerned with the circumstances under which information on aggregate prices, assumed to be common knowledge, is, or is not sufficient to reveal aggregate states.

Note that the aggregate budget constraint will hold at all times and in all states of nature irrespective of how individual consumption is determined and whatever the level of aggregate consumption. Even if this means linearised capital falls without bound, aggregate capital in levels is bounded below by zero.

If we define the state vector as $\xi_t = \begin{bmatrix} k_t & a_t \end{bmatrix}'$, we can rewrite (27) and (28), using (22), in the form of (7) as:

$$\xi_{t+1} = F_\xi \xi_t + F_c c_t + F_u u_{t+1} \quad (29)$$

³²We simplify the specification of the law of motion of capital and the measurement process by following Campbell's assumption in the first half of his paper that labour supply is fixed. This means that we have only one of the two potential forms of endogeneity in the filtering problem ($H_c = 0$ but $F_c \neq 0$). Given that we can exploit the "parallel problem" of Proposition 1 in solving the filtering problem, setting $H_c = 0$ has no qualitative impact on the results. Note that there is no source of extraneous noise.

where

$$F_\xi = \begin{bmatrix} \lambda_1 & \lambda_2 \\ 0 & \phi \end{bmatrix}; F_c = \begin{bmatrix} 1 - \lambda_1 - \lambda_2 \\ 0 \end{bmatrix}; F_u = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (30)$$

Using Campbell's (1994, p. 469) derivation of the linearisation constants from underlying structural parameters, we have $\lambda_2 = \alpha(r + \delta) / ((1 + g)(1 - \alpha)) > 0$, where r is the steady state value of the return on capital, g is the growth rate, δ is the depreciation rate, and α is the exponent on labour in a Cobb-Douglas production function. The eigenvalues of F_ξ are $\phi \in [0, 1]$, and $\lambda_1 = \frac{1+r}{1+g}$. Under the standard assumptions of dynamic efficiency, therefore, $r > g \Rightarrow \lambda_1 > 1$ hence F_ξ , the autoregressive matrix of the states in the parallel problem of Proposition 1, has one explosive eigenvalue.

We now consider what information is available to agents in the model. The standard assumption that k_t and a_t can both be directly measured is arguably a questionable one³³. It appears more reasonable to assume that the information set is "market-consistent", that agents have information based on the prices in the markets they trade in. Using this assumption, the vector of possible observable aggregate variables is:

$$\begin{bmatrix} w_t \\ r_t \end{bmatrix} = \begin{bmatrix} 1 - \alpha & \alpha \\ -\lambda_3 & \lambda_3 \end{bmatrix} \begin{bmatrix} k_t \\ a_t \end{bmatrix} \quad (31)$$

,where w_t is the aggregate wage, r_t is the aggregate return on capital, and $\lambda_3 > 0$ is Campbell's third linearisation parameter. The first line of (31) is the marginal product relation for labour, the second uses Campbell's linearisation of the return on capital.

We now consider the implications of three possible information sets. The first assumes that both possible measured variables are observable. The other two are censored versions of the information set, in which there is only a single observable variable.

The apparently arbitrary nature of these censored information sets can be justified on several grounds. First, they provide a technical illustration that contrasts two simple cases in which asymptotic invertibility does and does not hold, which we show relates to the nature of the measurement process, and its interaction with the evolution of the states. Second, they are of some independent interest, since in any solution that *does* asymptotically replicate full information the pre-determined nature of capital in this model means that innovations to the two measured variables

³³For a discussion of this assumption, see, eg, Bomfim (2004), Porapakarm and Young (2008) and Graham and Wright (2010).

will be perfectly correlated. The examples therefore allow us to identify which of the two observable variables will be informationally redundant if full information can be replicated. Third, as we discuss in section 6.4, each restricted information set can be derived as a special case of fully specified models.

6.1 An instantaneously invertible information set

In the first case we assume that $i_t = \begin{bmatrix} w_t & r_t \end{bmatrix}'$. Given this information set, we have $n = r = 2$ and by inspection of (31) the conditions for instantaneous invertibility in Definition 2 are satisfied. Thus factor prices provide sufficient information to infer the state vector from t -dated information (and hence the Kalman filter is redundant).

6.2 An asymptotically invertible information set

We first consider $i_t = w_t$, hence $H' = \begin{bmatrix} 1 - \alpha & \alpha \end{bmatrix}$. With a single shock, we thus have $n = s = 1$, hence the first condition in Proposition 3 is satisfied. The second is also satisfied, since $(H'_\xi F_u)^{-1} = \left(\begin{bmatrix} 1 - \alpha & \alpha \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix}' \right)^{-1} = \alpha^{-1} \neq 0$. For the third condition, we have, using the formulae in the Proposition,

$$\begin{aligned} \tilde{\beta}(Q) &= F_u (H'_\xi F_u)^{-1} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \frac{1}{\alpha} = \begin{bmatrix} 0 \\ \frac{1}{\alpha} \end{bmatrix} \\ \left(I - \tilde{\beta}(Q) H'_\xi \right) F_\xi &= \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 \\ \frac{1}{\alpha} \end{bmatrix} \begin{bmatrix} 1 - \alpha & \alpha \end{bmatrix} \right) \begin{bmatrix} \lambda_1 & \lambda_2 \\ 0 & \phi \end{bmatrix} \\ \Lambda \left(\left(I - \tilde{\beta}(Q) H'_\xi \right) F_\xi \right) &= \lambda_1 - \left(\frac{1 - \alpha}{\alpha} \right) \lambda_2 \end{aligned} \quad (32)$$

and thus, again, using Campbell (1994) the single critical eigenvalue³⁴ is, in terms of structural parameters,

$$\Lambda = \lambda_1 - \left(\frac{1 - \alpha}{\alpha} \right) \lambda_2 = \frac{1 - \delta}{1 + g} < 1$$

Hence the fixed point is stable, or, equivalently, using Corollary 2, filtering errors converge to zero in the neighbourhood of the full information equilibrium. Since it can also be shown that the model satisfies the conditions in Proposition 2, it follows that this is the unique steady state of the Kalman filter.

³⁴Recall that, from Corollary 2, the autoregressive matrix of filtering errors will have only $r - n = 1$ non-zero eigenvalue.

Thus the information set is asymptotically invertible: full information can be replicated from a sufficiently long history of the aggregate wage alone, and hence the history of returns is informationally redundant.

6.3 A non-invertible information set

We now consider the alternative censored information set, $i_t = r_t$, hence $H' = \begin{bmatrix} -\lambda_3 & \lambda_3 \end{bmatrix}$. The first two conditions in 3 are again trivially satisfied, with $n = s = 1$; and $(H'_\xi F_u)^{-1} = \left| \begin{bmatrix} -\lambda_3 & \lambda_3 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix}' \right| = \lambda_3^{-1} \neq 0$. Thus full information is again a fixed point of the Kalman filter given this information set. However the third condition is not satisfied: it is not a stable fixed point. To see this, we again use the formulae in the Proposition to give

$$\begin{aligned} \tilde{\beta}(Q) &= F_u (H'_\xi F_u)^{-1} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \frac{1}{\lambda_3} = \begin{bmatrix} 0 \\ \frac{1}{\lambda_3} \end{bmatrix} \\ (I - \tilde{\beta}(Q) H'_\xi) F_\xi &= \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 \\ \frac{1}{\lambda_3} \end{bmatrix} \begin{bmatrix} -\lambda_3 & \lambda_3 \end{bmatrix} \right) \begin{bmatrix} \lambda_1 & \lambda_2 \\ 0 & \phi \end{bmatrix} \\ \Lambda \left((I - \tilde{\beta}(Q) H'_\xi) F_\xi \right) &= \Lambda \left(\begin{bmatrix} \lambda_1 & \lambda_2 \\ \lambda_1 & \lambda_2 \end{bmatrix} \right) = \lambda_1 + \lambda_2 > 1 \end{aligned} \quad (33)$$

Hence this restricted information set is non-invertible.

To see the intuition for this result, and for the contrast with the previous example, recall that, from the representation in (25), the matrices in (32) and (33) are both the autoregressive matrices of the filtering error f_t in the neighbourhood of an equilibrium that replicates full information. In the information set with only the history of wages filtering errors are stationary, hence any initial error in estimating the state decays back to zero. When the information set consists only of the history of the return on capital, filtering errors are explosive in the neighbourhood of full information and the history of the aggregate wage, however poorly measured, is never informationally redundant.

For this example it can again be shown that the information set *does* satisfy the two conditions in Proposition 2 for the existence of a unique steady state Kalman filter; but the non-invertibility of the information set rules out the possibility that this can replicate full information. In Section 7.2 we discuss the qualitative nature of the equilibrium for the economy that arises from this non-invertible information set.

The solution to the filtering problem in this third example implies that, using

Proposition 4a the single observable variable r_t has a fundamental ARMA(2,1) representation. The order of this representation, and its autoregressive roots, are identical to those under full information (and hence to the two preceding invertible cases). The moving average roots will however differ, as will the innovation ε_t . In the invertible cases this will simply be a scaling of the productivity shock, u_t ; whereas in this case it will be a composite of current and lagged values of u_t . However, since $n = s$, Proposition 4c implies that there will be an alternative nonbasic nonfundamental ARMA(3,2) representation of r_t , the innovation to which will simply be a scaling of u_t .³⁵

6.4 Discussion

Although we have discussed the three cases purely in terms of the aggregate economy, each of the three examples can be related to an underlying economy with many agents. Note that, given symmetry of information sets on the aggregate economy, it is the nature of the parallel problem that allows us to focus on aggregates, since the solution to the filtering problem is independent of exactly how any particular agent determines their consumption.

The first example can be derived from an economy with many identical agents earning the same wage and the same return. The example shows that even if these agents cannot measure the aggregate states directly, the information in the market prices they observe is sufficient to recover the states. Graham and Wright (2010) show that the same information set, defined in terms of the *average* wage and the return on capital, will also be common knowledge in a heterogeneous economy with complete markets.

The second example shows that, given a sufficiently long history of the economy, market-consistent information sets reveal *more* information than is actually required to replicate full information, since in the limit the history of the return on capital becomes informationally redundant. This example can be seen as the limiting case of a model with noise in measuring aggregate returns on capital, due for example, to market frictions or noise traders in financial markets, and shows that such noise will have no impact on the equilibrium of the stochastic growth model, as long as the history of the aggregate wage is observable. The assumption of frictionless and complete financial markets is thus not of itself a necessary condition for replication of full information.

³⁵The AR roots in the fundamental representation will be the eigenvalues of G , hence will depend on the nature of the consumption function. The additional autoregressive root in the nonfundamental representation will be the autoregressive parameter of the single filtering error, $f_t = k_t - \hat{k}_t$.

However, some forms of market incompleteness *can* imply that the market-consistent information set is non-invertible. Our third example can be derived from the framework of Graham and Wright (2010), who analyse an incomplete markets version of the stochastic growth model in which there is a unified market for capital, but households earn a wage with an idiosyncratic component. As a result their (agent-specific) market-consistent information set does not contain the history of the aggregate wage.³⁶ In a limiting case in which each household’s individual wage is an arbitrarily poor signal of the aggregate wage, it is possible to show that the (common) information set on the aggregate economy consists only of the history of the return on capital, and hence is non-invertible, as in our third example.³⁷

7 Extensions and Applications

7.1 Data Vintages

A common assumption (see for example Collard and Dellas, 2006) is that data quality may improve over time, such that lagged values of state variables may be measured with less error than current values. There is also some evidence for this as an empirical phenomenon with published data (see, e.g., Orphanides and Van Norden, 2002).

Data vintages can be incorporated into our framework by extending the state space to include lags of the underlying states, with any associated measurement error in the relevant line of the measurement equation having lower variance than in the line relating to current states. However, this begs the question of how any such improvement in data quality may arise. One answer may be that statistical offices effectively engage in an informal version of “backward-smoothing”, whereby the Kalman filter can be used to derive improved estimates of underlying states by working backwards in time (see for example, Harvey, 1989, Section 3.6) and thus exploiting the benefits of hindsight. But, to the extent that this is the explanation, it clearly should have no impact on forward-looking behaviour at all since (at best) it implies that later vintages of data simply exploit the same information set that is used in deriving the best estimates of the current state variables and thus cannot improve the accuracy of these estimates.

Some form of backward-smoothing by statistical offices may also mean that

³⁶We discuss the properties of this economy further below, in Section 7.2.

³⁷It is also possible to show that in this limiting case there is a modified law of motion of aggregate capital; but it is still of the same general form as (27). Furthermore, aggregate consumption can be written as the optimising decision of a single representative agent. Details of the derivation of this limiting case are available from the authors.

econometric impulse responses estimated using historic data may be closer to true impulse responses (i.e., including the impact of the filtering process), but, as discussed in Section 4.3 this does not mean that such impulse responses would be observable in real time.

7.2 Heterogeneous information

Our framework has considered a class of models where the signal extraction problem is either that of a single optimising agent, or a set of such agents who share the same information set. The two types of problem are mathematically identical, since, if, for example, agents A and B both know the structural model and have the same information set, agent A does not need to *observe* Agent B’s behaviour: A can simply infer it from the common information set and structural properties.³⁸

There is potentially a much wider class of models in which individual agents, or types of agent, have overlapping, but not common information sets. In such models the techniques outlined in this paper are still a crucial part of the solution when states are endogenous; and limiting cases can have an identical form to the symmetric information equilibrium set out in this paper.

Graham and Wright (2010), for example, solve an incomplete markets stochastic growth model with a large number of heterogeneous households. In this economy each household faces a symmetric filtering problem of the same general form as (7) and (8). Each household has an idiosyncratic capital stock which is endogenous to their own consumption and labour supply choices, and faces idiosyncratic labour supply shocks; all other elements of the state vector are aggregate, and hence exogenous, and common to each agent’s filtering problem. Each household’s filtering problem is solved as a parallel problem as in Proposition 1, taking the dynamics of the aggregate economy as given.

The attraction of this approach is that the filtering problem arises from the structure of the economy, rather than from any extraneous noise or an assumption of exogenous informational restrictions. But a significant complication is that, if all households *know* that all other households are solving the same problem, each household also knows that the dynamics of the aggregate economy must be affected (via the aggregate capital stock) by the dynamic choices of consumers in aggregate. This aggregate behaviour is unobservable to any individual household; but households can infer the nature of the choices (and hence the dynamics of the aggregate economy) from their own behaviour. Graham and Wright show that the resulting solution requires each household to form estimates of a “hierarchy of average

³⁸This is the basis for the solution method used by Svensson and Woodford (2004).

expectations“ (Townsend, 1983, Woodford, 2003, Nimark, 2007) of the underlying non-expectational states, resulting in an infinite dimensional state vector. However, extending new techniques developed by Nimark (2008), the model can be solved, to an arbitrary degree of precision, with a finite state vector.

In this economy optimal behaviour for each household remains, as in this paper, certainty-equivalent with respect to the full state vector (including the hierarchy of expectations), but is not, in general, certainty-equivalent with respect to the underlying non-expectational states. Thus in contrast to the solution here, the aggregate economy cannot in general be given an isomorphic representation in terms of a notional full information economy with a different set of innovations, as in Corollary 1. However, if there is sufficient heterogeneity, there is a limiting case that is certainty-equivalent, and which therefore does have such a representation. Furthermore, on available evidence on the relative variance of idiosyncratic versus aggregate shocks, the calibrated economy is very close to this limiting case. Thus certainty equivalence appears to provide at least a reasonable approximation even of an economy with both structural and informational heterogeneity. The filtering problem in this limiting case is of the same form as that of the benchmark stochastic growth model with an artificially censored non-invertible information set that we analysed in Section 6.3.

Graham and Wright (2010) show that in this framework the steady-state Kalman filter implies a positive aggregate productivity shock is interpreted on impact largely as indicating that capital has previously been over-estimated, which therefore causes aggregate consumption to fall, rather than rise, as it would under full information. Given the closeness of the economy to the limiting case of extreme heterogeneity discussed above, impulse responses can be well approximated by certainty-equivalent responses to state estimates, as in (13); or equivalently by responses in an isomorphic full information (and in this context, complete markets) economy (as in Corollary 1) that is regularly hit by pseudo-shocks to capital.

8 Conclusions

In this paper we have presented a general method of solving the signal extraction problem in linearised DSGE models, which allows for endogeneity between dynamic choice variables and both measured and state variables. We derive a number of key features of such economies, that relate to the nature of the solution of the Kalman filter, focussing in particular on whether the information set is invertible, i.e., whether it can replicate full information, at least asymptotically. We have

emphasised the application of our techniques to DSGE models. However most of our results are quite general, and in principle applicable in a wide variety of contexts where dynamic optimisation problems involve states that are both endogenous, and not directly observable. As such the techniques set out in this paper broaden out further the already wide scope for application of the Kalman filter.

References

- Anderson, D and John B. Moore, (1979), "Optimal Filtering", Dover
- Blanchard, Olivier and James A Kahn, (1980), "The Solution of Linear Difference Models under Rational Expectations", *Econometrica* 48 (5), pp.1305-1312.
- Bomfim, Antulio N, (2004), "Measurement error in general equilibrium: the aggregate effects of noisy economic indicators", *Journal of Monetary Economics* 48, pp.505-603.
- Campbell, John Y, (1994), "Inspecting the mechanism: an analytical approach to the stochastic growth model", *Journal of Monetary Economics* 33, pp.463-506.
- Collard, Fabrice and Harris Dellas, (2006), "Imperfect Information and Inflation Dynamics", Working Paper.
- Fernandez-Villaverde, J, Rubio-Ramirez, F, Sargent, T and Watson, M (2007) "ABCs (and Ds) of Understanding VARs", *American Economic Review*, pp 1021-1026
- Graham, Liam and Stephen Wright, (2010), "Information, heterogeneity and market incompleteness", *Journal of Monetary Economics*, 57, pp 164 - 174.
- Hansen, L and Sargent, T (2005), "Recursive Models of Dynamic Linear Economies", monograph
- Harvey, Andrew (1981) "Time Series Models" Philip Allan, London
- Harvey, Andrew (1989) "Forecasting, Structural Time Series Models and the Kalman Filter" Cambridge University Press
- Hamilton, James, (1994), "Time Series Analysis", Princeton University Press.

- Keen, Benjamin D, (2004), "The Signal Extraction Problem Revisited:Its Impact on a Model of Monetary Policy", Working Paper.
- Lippi, Marco and Reichlin, Lucrezia (1994) "VAR analysis, nonfundamental representations, Blaschke matrices", *Journal of Econometrics*, 63 pp 307-325
- Lorenzoni, Guido, (2010), "A Theory of Demand Shocks". *American Economic Review*, forthcoming
- McCallum, Bennett T, (1998), "Solutions to linear rational expectations models: a compact exposition", *Economics Letters* 61, pp.143-147.
- Mehra, Rajnish and Edward C Prescott, (1980), "Recursive Competitive Equilibrium: The Case of Homogeneous Households", *Econometrica* 48 (6), pp.1365-1379.
- Muth, John F, (1961), "Rational Expectations and the Theory of Price Movements", *Econometrica* 29 (3), pp.315-335.
- Nimark, Kristoffer, (2007), "Dynamic Higher Order Expectations", Working Paper
- Nimark, Kristoffer, (2008), "Dynamic Pricing and Imperfect Common Knowledge". *Journal of Monetary Economics*, 55 (2008), pp365-382.
- Orphanides, Athanasios and Simon Van Norden (2002) "The unreliability of output-gap estimates in real time", *Review of Economics and Statistics*, 84, 4, pp 569-583
- Pearlman, Joseph, David Currie, and Paul Levine,(1986), "Rational Expectations Models with Partial Information", *Economic Modeling* 3, pp 90-105.
- Pearlman, Joseph, (1992), "Reputational and nonreputational policies under partial information", *Journal of Economic Dynamics and Control* 16 (2), pp.339-357.
- Porapakarm, Ponpoje and Eric R. Young (2008), "Information Heterogeneity in the Macroeconomy", Working paper
- Robertson, D and Wright, S (2010) "Bounds for the Predictive R-Squared", working paper.
- Söderström, T (1994) "Discrete-Time Stochastic Systems" Prentice Hall, London, New York.
- Svensson, Lars EO and Michael Woodford, (2004), "Indicator variables for optimal policy under asymmetric information.", *Journal of Economic Dynamics and Control* 28 (4), pp.661-690.

Svensson, Lars EO and Michael Woodford, (2003), "Indicator variables and optimal policy", *Journal of Monetary Economics* 50 (3), pp.691-720.

Townsend, Robert M, (1983), "Forecasting the Forecasts of Others", *Journal of Political Economy* 91 (4), pp 546-588.

Whittle, P (1983) "Prediction and Regulation by Least-Squares Methods", University of Minnesota Press

Woodford, M, (2003), "Imperfect Common Knowledge and the Effects of Monetary Policy," in P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford, eds., *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honour of Edmund S. Phelps*, Princeton:Princeton University Press.

Table 1. Time Series Properties of Observables

| Nature of Information Set | Number of Non-Zero Eigenvalues of State Representation | | | VARMA(n,p,q) order of observables, i_t (p =autoregressive order, q = moving average order) | | | |
|---|--|-----------------------------|----------------------------------|---|----------|--------------------|-------|
| | Full Information, equation (10) | True Process, equation (25) | Estimated Process, equation (19) | True Process | | Observable Process | |
| | | | | p | q | p | q |
| Instantaneously Invertible | r | r | r | 1 | 0 | 1 | 0 |
| Asymptotically Invertible | r | r | r | $r-n+1$ | $r-n$ | $r-n+1$ | $r-n$ |
| Non-invertible, exogenous state process ($F_c=0$) | r | r | r | $r-n+1$ | $r-n$ | $r-n+1$ | $r-n$ |
| Non-invertible, dynamically endogenous state process ($F_c \neq 0$) | r | $2r - n$ | r | $2(r-n)+1$ | $2(r-n)$ | $r-n+1$ | $r-n$ |

Appendix

A Derivation of Equations (7) and (8)

We first stack equations (2) on top of (3) and (6) to derive the law of motion for the state variables, $\xi_t = \begin{bmatrix} k_t & z_t & w_t \end{bmatrix}'$, and respecify (5) accordingly,

$$\xi_{t+1} = D_{\xi\xi}\xi_t + D_{\xi y}y_t + v_{t+1} \quad (\text{A.1})$$

$$i_t = D_{i\xi}\xi_t + D_{iy}y_t \quad (\text{A.2})$$

where

$$D_{\xi\xi} = \begin{bmatrix} B_{kk} & B_{kz} & 0 \\ 0 & B_{zz} & 0 \\ 0 & 0 & B_{ww} \end{bmatrix}; \quad D_{\xi y} = \begin{bmatrix} B_{ky} \\ 0 \\ 0 \end{bmatrix}; \quad D_{i\xi} = \begin{bmatrix} C_{ik} & C_{iz} & C_{iw} \end{bmatrix}; \quad D_{iy} = C_{iy};$$

$$Q = E(v_t v_t') = E(v_t v_t') = \begin{bmatrix} 0_{r_k \times r_k} & 0_{r_k \times s} \\ 0_{s \times r_k} & S \end{bmatrix}; \quad S = \begin{bmatrix} S_{\zeta\zeta} & S_{\zeta\omega} \\ S'_{\zeta\omega} & S_{\omega\omega} \end{bmatrix}$$

We next partition y_t as $y_t = \begin{bmatrix} c_t' & x_t' \end{bmatrix}'$ and express (1) conformably as:¹

$$\begin{bmatrix} A_{cc} & A_{cx} \\ 0 & 0 \end{bmatrix} E_t \begin{bmatrix} c_{t+1} \\ x_{t+1} \end{bmatrix} = \begin{bmatrix} B_{cc} & B_{cx} \\ B_{xc} & B_{xx} \end{bmatrix} \begin{bmatrix} c_t \\ x_t \end{bmatrix} + \begin{bmatrix} B_{c\xi} \\ B_{x\xi} \end{bmatrix} \xi_t \quad (\text{A.3})$$

where the first block of equations are expectational difference equations and thus represent dynamic choice variables such as consumption or policy variables. The second block of equations represent purely static relationships (for example, intratemporal optimality conditions, production functions, identities, etc) that can be substituted out. Using these, assuming B_{xx}^{-1} exists² we can substitute out using

$$x_t = -B_{xx}^{-1} [B_{xc}c_t + B_{x\xi}\xi_t] = D_{xc}c_t + D_{x\xi}\xi_t \quad (\text{A.4})$$

and write the state and measurement equation in their final form in the main text

¹This form for A_{yy} will usually follow naturally from the structure of the model, but as long as A_{yy} is singular (if it is not, then the vector x_t will be empty) this structure can always be achieved by an appropriate linear re-weighting of the elements of y_t . The sub-matrix A_{cx} may also in principle contain columns of zeros.

²The case where B_{xx}^{-1} does not exist implies that some elements of x_t can be expressed as linear combinations of other elements, and can thus be trivially dealt with by substitution.

as

$$\xi_{t+1} = F_\xi \xi_t + F_c c_t + v_{t+1} \quad (\text{A.5})$$

$$i_t = H'_\xi \xi_t + H_c c_t \quad (\text{A.6})$$

where $D_{\xi y} = \begin{bmatrix} D_{\xi c} & D_{\xi x} \end{bmatrix}$, $D_{iy} = \begin{bmatrix} D_{ic} & D_{ix} \end{bmatrix}$, etc, and

$$F_\xi = D_{\xi\xi} + D_{\xi x} D_{x\xi}; \quad F_c = D_{\xi c} + D_{\xi x} D_{xc}; \quad (\text{A.7})$$

$$H'_\xi = D_{i\xi} + D_{ix} D_{x\xi} \quad H_c = D_{ic} + D_{ix} D_{xc} \quad (\text{A.8})$$

Note that the substitutions involved in deriving (7) and (8) are by no means innocuous in informational terms.

First, even static relationships may require informational assumptions. Since they may involve linear combinations of state variables it may be of considerable importance whether these combinations, or the elements of x_t themselves, are in the information set I_t . The form of the measurement equation allows for the possibility that elements of x_t may be observable, whether directly or indirectly, but there are interesting cases where they are not.³

The nature of the expectational difference equations satisfied by c_t , the dynamic choice variables, may also have important informational implications. While this framework can in principle accommodate any structure to the top block of equations in (A.3), certain structures may require assumptions about the nature of the information set. Thus if we substitute out for x_t using (A.4) and (7) and use (13) we can write the top block, applying the law of iterated expectations, as

$$\{A_{cc}\eta' + A_{cx}(D_{xc}\eta' + D_{x\xi})\} G\widehat{\xi}_t = \{B_{cc} + B_{cx}D_{xc}\}\eta'\widehat{\xi}_t + \{B_{c\xi} + B_{cx}D_{x\xi}\}\xi_t$$

which depends on ξ_t as well as $\widehat{\xi}_t$. For such a formulation to be informationally feasible in this precise form, the linear combination of states given by $\{B_{c\xi} + B_{cx}D_{x\xi}\}\xi_t$ must be observable, and therefore should also be an element of i_t . In principle this may significantly alter the information set and hence the nature of the filtering problem (although the rationale for this combination being observable should be justifiable). However, it will not alter the certainty-equivalent nature of the consumption function. If this linear combination is indeed observable, then (from Corollary 4) efficiency of the state estimates requires that they satisfy the adding up constraint $\{B_{c\xi} + B_{cx}D_{x\xi}\}\xi_t = \{B_{c\xi} + B_{cx}D_{x\xi}\}\widehat{\xi}_t$ thus allowing the top block to be written

³In Graham and Wright (2009a), for example, the aggregate wage is an element of x_t that is not observable, given heterogeneous labour and incomplete markets.

entirely in terms of state estimates, as

$$\{A_{cc}\eta' + A_{cx}(D_{xc}\eta' + D_{x\xi})\}G\widehat{\xi}_t = [\{B_{cc} + B_{cx}D_{xc}\}\eta' + B_{c\xi} + B_{cx}D_{x\xi}]\widehat{\xi}_t$$

which results in an undetermined coefficients problem identical to that under full information. Note also that the nature of the undetermined coefficients problem is unchanged if this linear combination of states is not observable, but is replaced by the same combination of state estimates.

This issue does not, of course, arise if, as in many contexts (for example consumption Euler equations) B_{cx} and $B_{c\xi}$ are zero.

B Proof of Proposition 1

For completeness, and because some elements of the proof are of relevance to later analysis, we present a full proof of the Proposition. Most strictly algebraic elements of the proof are however to be found in the existing literature on optimal control. For example, Söderström (1994, Ch. 6) derives the solution to a filtering problem with what we refer to here as “dynamic endogeneity” (i.e. $F_c \neq 0$). His solution is identical to ours if we set $H_c = 0$, once we allow for differences in notation; he does not however explicitly draw out the link with the “parallel problem”. We also show in Section B.3 of this appendix that our approach yields identical answers to those derived in Svensson and Woodford (2003), despite their rather unusual statement of the filtering process.

We assume that in some period $t - 1$ initial estimates of the states ξ_t and P_t are available, that must satisfy $E_{t-1}\widehat{\xi}_t = E_{t-1}\xi_t$ by the law of iterated expectations, given the definition of $\widehat{\xi}_t$. This condition will always be satisfied if, at $t = 0$, $E_0\widehat{\xi}_1 = E_0\xi_1$.

B.1 Forecasting i_t

Using (8)

$$E_{t-1}i_t = H'_\xi E_{t-1}\xi_t + H_c\eta' E_{t-1}\widehat{\xi}_t \quad (\text{B.1})$$

where the second line follows by the law of iterated expectations. The error of this forecast is, using (20)

$$\varepsilon_t \equiv i_t - E_{t-1}i_t = H'_\xi \left[\xi_t - E_{t-1}\widehat{\xi}_t \right] + H_c\eta' \left[\widehat{\xi}_t - E_{t-1}\xi_t \right] \quad (\text{B.2})$$

We then treat (19), the process for the estimated states, as a conjectured solution to the filtering process, which we show below be verified by the actual solution.

Conditional upon this conjectured solution

$$\varepsilon_t = J_t' [\xi_t - E_{t-1}\xi_t] \quad (\text{B.3})$$

where

$$J_t' = [I_n - H_c \eta' \beta_t]^{-1} H_c' \quad (\text{B.4})$$

Thus we have, using (B.3) and (15)

$$E[\varepsilon_t \varepsilon_t'] = J_t' E[(\xi_t - E_{t-1}\xi_t)(\xi_t - E_{t-1}\xi_t)'] J_t = J_t' P_t J_t \quad (\text{B.5})$$

B.2 Deriving β_t , M_t and P_t .

Since (conditional upon β_t and hence J_t) innovations to i_t depend only on unobservable errors in forecasting the states, the Kalman Gain matrix, β_t in the updating equation (19) is

$$\beta_t = \{E[(\xi_t - E_{t-1}\xi_t) \varepsilon_t']\} \{E[\varepsilon_t \varepsilon_t']\}^{-1} \quad (\text{B.6})$$

hence, using (B.3), (15), and (B.5),

$$\beta_t = P_t J_t [J_t' P_t J_t]^{-1} \quad (\text{B.7})$$

and the MSE of the state estimates can be written as

$$\begin{aligned} M_t &= E\left[\left(\xi_t - E_{t-1}\widehat{\xi}_t\right)\left(\xi_t - E_{t-1}\widehat{\xi}_t\right)'\right] - \beta_t E\left[\varepsilon_t\left(\xi_t - E_{t-1}\widehat{\xi}_t\right)'\right] \\ &= P_t - \beta_t J_t' P_t = [I_{r+n} - \beta_t J_t'] P_t \end{aligned} \quad (\text{B.8})$$

however these do not yet constitute closed form solutions since, via (B.4), J_t depends on β_t .

To show that the recursion for P_{t+1} is F_c -independent, note that the forecast error in predicting the states in period $t+1$ is, using (11), and (7),

$$\begin{aligned} \xi_{t+1} - E_t \widehat{\xi}_{t+1} &= F_\xi \xi_t + F_c \eta' \widehat{\xi}_t + v_{t+1} - (F_\xi + F_c \eta') \widehat{\xi}_t \\ &= F_\xi (\xi_t - \widehat{\xi}_t) + v_{t+1} \end{aligned} \quad (\text{B.9})$$

and is thus independent of F_c . Hence, using the orthogonality assumptions and (14),

$$P_{t+1} = F_\xi M_t F_\xi' + Q. \quad (\text{B.10})$$

To show that P_{t+1} and M_t are H_c -independent, define

$$J'_t = K_t^{-1} H'_\xi, \quad (\text{B.11})$$

where K_t is the (as yet unknown) matrix that satisfies

$$K_t = (I_n - H_c \eta' \beta_t). \quad (\text{B.12})$$

$$\Rightarrow (J'_t P_t J_t)^{-1} = \left(K_t^{-1} H'_\xi P_t H_\xi (K_t^{-1})' \right)^{-1} = K_t' (H'_\xi P_t H_\xi)^{-1} K_t \quad (\text{B.13})$$

and hence, substituting (B.11) and (B.13) into (B.7), we obtain

$$\beta_t = P_t H_\xi (H'_\xi P_t H_\xi)^{-1} K_t \quad (\text{B.14})$$

and thus

$$\begin{aligned} \beta_t J'_t &= H_\xi (K_t^{-1})' K_t' (H'_\xi P_t H_\xi)^{-1} K_t K_t^{-1} H'_\xi \\ &= H_\xi (H'_\xi P_t H_\xi)^{-1} H'_\xi. \end{aligned} \quad (\text{B.15})$$

We thus have

$$M_t = \left(I_r - P_t H_\xi (H'_\xi P_t H_\xi)^{-1} H'_\xi \right) P_t, \quad (\text{B.16})$$

$$P_{t+1} = F_\xi \left(I_r - P_t H_\xi (H'_\xi P_t H_\xi)^{-1} H'_\xi \right) P_t F'_\xi + Q, \quad (\text{B.17})$$

Thus the recursions for M_t and P_t do not depend on F_c or K_t (and hence H_c) and can thus be derived by setting $H_c = F_c = 0$ as in the parallel problem. If we define $\tilde{\beta}_t$ as in (18) then the above formulae are identical to (16) and (17) in Proposition 1. We also have, using (B.15)

$$\beta_t J'_t = \tilde{\beta}_t H'_\xi \quad (\text{B.18})$$

Finally we need to obtain an expression for J_t itself, and hence for β_t . Equations (B.4) and (B.7) imply the seemingly nonlinear equation

$$J'_t = \left(I_n - H_c \eta' P_t J_t (J_t P_t J_t^{-1}) \right)^{-1} H'_\xi.$$

However, using (B.11) and (B.14), we obtain

$$K_t^{-1} H'_\xi = J'_t = \left(I_n - H_c \eta' P_t H_\xi (H'_\xi P_t H_\xi)^{-1} K_t \right)^{-1} H'_\xi, \quad (\text{B.19})$$

$$\Rightarrow K_t^{-1} H'_\xi - H_c \eta' P_t H_\xi (H'_\xi P_t H_\xi)^{-1} H'_\xi = H'_\xi. \quad (\text{B.20})$$

Recalling (B.11) once again, we find

$$\begin{aligned} J'_t &= H'_\xi + H_c \eta' P_t H_\xi (H'_\xi P_t H_\xi)^{-1} H'_\xi \\ &= \left(I_n + H_c \eta' P_t H_\xi (H'_\xi P_t H_\xi)^{-1} \right) H'_\xi \end{aligned} \quad (\text{B.21})$$

$$\Rightarrow K_t = \left[I_n + H_c \eta' P_t H_\xi (H'_\xi P_t H_\xi)^{-1} \right]^{-1} \quad (\text{B.22})$$

Using (18), these can be expressed as

$$J'_t = \left(I_n + H_c \eta' \tilde{\beta}_t \right) H'_\xi \quad (\text{B.23})$$

$$K_t = \left(I_n + H_c \eta' \tilde{\beta}_t \right)^{-1} \quad (\text{B.24})$$

which, after substituting from (B.24) into (B.14) gives (21), completing the proof of Proposition 1. ■

B.3 Comparison with Svensson and Woodford (2003) and Pearlman et al (1986)

Svensson and Woodford (2003) have a structural model which in reduced form is extremely close to ours. Their equations (15) and (16) correspond directly to our state and measurement equations (7) and (8), after substituting from (13). Using their notation their equation (22) is

$$X_{t|t} = X_{t|t-1} + K \left[L (X_t - X_{t|t-1}) + v_t \right] \quad (\text{B.25})$$

where, $X_{t|t}$ in their notation corresponds to $\hat{\xi}_t$ in ours, and Z_t to our i_t . They then assert that this allows them to identify K as "(one form of) the Kalman Gain Matrix" (which they assume, without proof, will converge to a fixed matrix). However, by the usual convention in the literature the Kalman gain updates in response to a forecast error. The square bracketed expression is not a forecast error. Using their (16), the true forecast error in their framework is

$$Z_t - E_{t-1} Z_t = L (X_t - X_{t|t-1}) + M (X_{t|t} - X_{t|t-1}) + v_t \quad (\text{B.26})$$

where the endogeneity of the measured variables to the response of the estimated states is evident.

However, it turns out that, despite the somewhat unusual basis for their derivation, their final result is in fact identical to our own. If we re-express their (22) in

our own notation (apart from the matrix K), it becomes

$$\widehat{\xi}_t - E_{t-1}\widehat{\xi}_t = K \left[H'_\xi \left(\xi_t - E_{t-1}\widehat{\xi}_t \right) \right] \quad (\text{B.27})$$

whereas we show that, in our notation, from (B.3), after substituting from the endogenous response of c_t , and assuming convergence, the updating rule in response to the forecast error in the measured variables is given by

$$\widehat{\xi}_t - E_{t-1}\widehat{\xi}_t = \beta \varepsilon_t = \beta J' \left(\xi_t - E_{t-1}\widehat{\xi}_t \right) \quad (\text{B.28})$$

But using their equations (24) and (25) (noting that we absorb the covariance matrix of measurement errors into Q , and hence P), their derivation implies, in our notation,

$$K = \widetilde{\beta}$$

thus in our notation K is identical to the Kalman gain matrix in the parallel, rather than the actual problem. But, from (B.18), we have $\beta J' = \widetilde{\beta} H'_\xi$, hence

$$\widehat{\xi}_t - E_{t-1}\widehat{\xi}_t = \beta J' \left(\xi_t - E_{t-1}\widehat{\xi}_t \right) = \widetilde{\beta} H'_\xi \left(\xi_t - E_{t-1}\widehat{\xi}_t \right) \quad (\text{B.29})$$

thus Svensson and Woodford's updating rule is in fact identical to our own. An equivalent updating rule is also given in Pearlman *et al* (1986) equation (39). However, neither of these papers note the equivalence of M_t and P_{t+1} in the parallel problem, nor do they derive convergence conditions.

C Proof of Proposition 2 and Corollary 2.

C.1 Proof of Proposition 2

Since β_t and M_t can both be expressed in terms of P_t and structural parameters a necessary and sufficient condition for convergence of all three matrices to a unique steady state is convergence of P_t to a unique steady state. Since P_t can be derived from the the parallel problem of Proposition 1 in which the states are exogenous we only need be concerned with the stability properties of that problem. Anderson and Moore (1979, pp 77-81) provide a proof of a unique stable steady state given controllability and detectability as defined in the main text for any invertible P_0 .

C.2 Proof of Corollary 2

We first restate (16), writing $F \equiv F_\xi$, $H \equiv H_\xi$ in this section, for brevity, as

$$P_{t+1} = F \left(I_r - P_t H (H' P_t H)^{-1} H' \right) P_t F' + Q. \quad (\text{C.1})$$

In other words, we are iterating the function $g : \mathbb{P}_r \rightarrow \mathbb{P}_r$, where \mathbb{P}_r denotes the set of all non-negative definite symmetric, real $r \times r$ matrices, and

$$g(P_t) = F \left(I_r - P_t H (H' P_t H)^{-1} H' \right) P_t F' + Q, \quad P_t \in \mathbb{P}_r. \quad (\text{C.2})$$

If the conditions set by Proposition 2 are satisfied, then this iteration is stable around a unique fixed point P

We first note a convenient simplification. Let

$$\widehat{F}(P_t) = F \left(I_r - P_t H (H' P_t H)^{-1} H' \right) = F \left(I_r - \widetilde{\beta}(P_t) H' \right) \quad (\text{C.3})$$

(where the second expression uses (18)) then:

Lemma 1 *The function $g : \mathbb{P}_r \rightarrow \mathbb{P}_r$ defined by (C.2) can be expressed, using (C.3), in the symmetric form*

$$g(P_t) = \widehat{F}(P_t) P_t \widehat{F}(P_t)' + Q, \quad (\text{C.4})$$

Proof. Using (C.3), we have

$$g(P_t) = \widehat{F}(P_t) P_t F' + Q \quad (\text{C.5})$$

$$\widehat{F} P_t \widehat{F}' = \widehat{F} P_t F' - F \left(I_r - \widetilde{\beta} H' \right) P_t H \widetilde{\beta}' F' \quad (\text{C.6})$$

but

$$\left(I_r - \widetilde{\beta} H' \right) P_t H \widetilde{\beta}' = P_t H \widetilde{\beta}' - \widetilde{\beta} H' P_t H \widetilde{\beta}' = P_t H \widetilde{\beta}' - P_t H (H' P_t H)^{-1} H' P_t H \widetilde{\beta}' = 0 \quad (\text{C.7})$$

■

As is usual in the analysis of fixed point iteration, we must calculate the (Fréchet) derivative of g at the fixed point P .

Lemma 2 *If $E \in \mathbb{P}_r$, then, letting $\widehat{F}_P = \widehat{F}(P)$*

$$g(P + E) = g(P) + \widehat{F}_P E \widehat{F}_P' + O(E^2). \quad (\text{C.8})$$

Thus if we let Dg_P denote the Fréchet derivative of the matrix function g at the point $P \in \mathbb{P}_r$, then

$$Dg_P(E) = \widehat{F}_P E \widehat{F}'_P \quad E \in \mathbb{P}_r. \quad (\text{C.9})$$

Proof. We have, using (C.2)

$$\begin{aligned} g(P + E) &= F(P + E)F' - F(P + E)H(H'PH + H'EH)^{-1}H'(P + E)F' + Q \\ &= F(P + E)F' - F(P + E)H \left[(H'PH)(I + (H'PH)^{-1}(H'EH)) \right]^{-1} H'(P + E)F' + Q \\ &= \dots - F(P + E)H \left[I - (H'PH)^{-1}(H'EH) \right] (H'PH)^{-1} H'(P + E)F' + Q + O(E^2) \end{aligned} \quad (\text{C.10})$$

$$= g(P) + \widehat{F}_P E \widehat{F}'_P + O(E^2). \quad (\text{C.11})$$

■

It is useful to restate (C.8) and (C.9) in Kronecker product notation, as

$$\text{vec}(g(P + E)) = \text{vec}(g(P)) + \widehat{F}_P \otimes \widehat{F}'_P \text{vec}(E) + O(\text{vec}(E^2))$$

hence, in this form the Frechet derivative is (using (C.3))

$$Dg_P = \widehat{F}_P \otimes \widehat{F}'_P = F \left(I_r - \widetilde{\beta}(P)H' \right) \otimes F \left(I_r - \widetilde{\beta}(P)H' \right) \quad (\text{C.12})$$

and thus as a corollary of Proposition 2, stability of the steady state implies that the matrix $F_\xi \left(I_r - \widetilde{\beta}(P)H'_\xi \right)$ must have eigenvalues with real parts strictly less than one in absolute value. Since products of matrices have common non-zero eigenvalues irrespective of order of multiplication this condition must also apply to the matrix $\left(I_r - \widetilde{\beta}(P)H'_\xi \right) F_\xi$. The proof of Corollary 4 shows that this matrix will be of rank $r - n$. ■

C.3 Proof of Corollary 1.

From the autoregressive representation of the estimated states in (19) it is evident that they have the same autoregressive form as the true states under full information in (10). To derive the implied innovation covariance matrix we have, using (20), (B.3), (B.5) and (B.18),

$$\begin{aligned} E \left[\left(\widehat{\xi}_{t+1} - E_t \widehat{\xi}_{t+1} \right) \left(\widehat{\xi}_{t+1} - E_t \widehat{\xi}_{t+1} \right)' \right] &= \beta E \left(\varepsilon_{t+1} \varepsilon'_{t+1} \right) \beta' \\ &= \beta J' P J \beta' = \widetilde{\beta} H'_\xi P H_\xi \widetilde{\beta}' \end{aligned} \quad (\text{C.13})$$

but hence, using (16), (18), (17), and exploiting symmetry of P and M

$$\begin{aligned}
E \left[\left(\widehat{\xi}_{t+1} - E_t \widehat{\xi}_{t+1} \right) \left(\widehat{\xi}_{t+1} - E_t \widehat{\xi}_{t+1} \right)' \right] &= PH_\xi [H'_\xi PH_\xi]^{-1} PH_\xi \widetilde{\beta}' \\
&= PH_\xi \widetilde{\beta}' = (\beta H' P)' \\
&= (P - M)' = Q + F_\xi M F'_\xi - M \quad (\blacksquare.14)
\end{aligned}$$

D Derivation of joint process for ξ_t and f_t in (25) and proofs of corollaries 3 to 5

D.1 Derivation of (25).

Using (7), (13) and (24) we have

$$\xi_{t+1} = F_\xi \xi_t + F_c \eta' \widehat{\xi}_t + v_{t+1} = (F_\xi + F_c \eta') \xi_t - F_c \eta' f_t + v_{t+1} \quad (\text{D.1})$$

$$= G \xi_t - F_c \eta' f_t + v_{t+1} \quad (\text{D.2})$$

For the estimated states we have, using (19), the definition of G in (10), (B.3) and (7)

$$\begin{aligned}
\widehat{\xi}_{t+1} &= G \widehat{\xi}_t + \beta \varepsilon_{t+1} = G \widehat{\xi}_t + \beta J' [\xi_{t+1} - G \widehat{\xi}_t] \\
&= G \widehat{\xi}_t + \beta J' [\xi_{t+1} - (F_\xi + F_c \eta') \widehat{\xi}_t] \quad (\text{D.3})
\end{aligned}$$

$$= G \widehat{\xi}_t + \beta J' [\xi_{t+1} - F_\xi \xi_t - F_c \eta' \widehat{\xi}_t + F_\xi (\xi_t - \widehat{\xi}_t)] \quad (\text{D.4})$$

$$= G \widehat{\xi}_t + \beta J' [v_{t+1} + F_\xi f_t] \quad (\text{D.5})$$

hence, using (11) and (B.18),

$$f_{t+1} = (G - F_c \eta' - \beta J' F_\xi) f_t + (I - \beta J') v_{t+1} \quad (\text{D.6})$$

$$= [I - \beta J'] F_\xi f_t + (I - \beta J') v_{t+1} \quad (\text{D.7})$$

$$= \left[I - \widetilde{\beta} H'_\xi \right] F_\xi f_t + \left(I - \widetilde{\beta} H'_\xi \right) v_{t+1}$$

Stacking (D.1) above (D.6) gives the joint process for ξ_{t+1} and f_{t+1} in (25).

D.2 Proof of Corollary 3

Since the filtering error process f_t is block recursive we can write the top block of equations as

$$\begin{aligned}\xi_{t+1} &= [I - GL]^{-1} v_{t+1} - [I - GL]^{-1} F_c \eta' f_t = \xi_{t+1}^* - [I - GL]^{-1} F_c \eta' f_t \\ &= \xi_{t+1}^* - [I - GL]^{-1} F_c \eta' [I - (I - \beta J') F_\xi L]^{-1} (I - \beta J') v_t\end{aligned}\quad (\text{D.8})$$

where ξ_{t+1}^* , the full information state process, is as given by (10), and the last line uses the lower block of (25). The incomplete information states are thus equal to the full information states plus a lag polynomial in the filtering error (itself a lag polynomial in the underlying shocks). Since the filtering error is stationary (from Corollary 2) and the full information process is non-explosive (from Assumptions 1 and 2) the incomplete information process is also non-explosive. Since there may be permanent productivity or other shocks, G may have unit eigenvalues, implying permanent effects of these shocks. But permanent effects will only arise with respect to rows of v_t for which the relevant rows of F_c are, by Assumption 1, zero (the shock processes are exogenous). Hence filtering error will only cause transitory deviations from the full information outcome, proving the Corollary. ■

D.3 Proof of Corollary 4

Using (18) we have

$$H'_\xi \tilde{\beta} = H'_\xi P H_\xi (H'_\xi P H_\xi)^{-1} = I_n \quad (\text{D.9})$$

If we pre-multiply (D.6) by H'_ξ and use (D.9) we have

$$H'_\xi f_{t+1} = H'_\xi \left[I - \tilde{\beta} H'_\xi \right] F_\xi f_t + H'_\xi \left(I - \tilde{\beta} H'_\xi \right) v_{t+1} = 0 \quad (\text{D.10})$$

thus proving Corollary 4, and at the same time showing that the matrix $\left[I - \tilde{\beta} H'_\xi \right] F_\xi$ is of rank $\leq r - n$, completing the proof of Corollary 2. ■

D.4 Proof of Corollary 5

By inspection of (25), if F_ξ has explosive eigenvalues in its sub-matrix F_{kk} and $\beta_k = 0$, then the matrix $\left(I - \tilde{\beta} H' \right) F_\xi$ will also have explosive eigenvalues, which, from Corollary 2, contradicts stability of the recursion for P_{t+1} , thus proving Corollary 5. ■

E Proof of Proposition 3

E.1 Proof of Proposition 3

Using (B.6) our original definition of β_t , and (20), we have

$$\beta_t = E \left((\xi_{t+1} - E_t \xi_{t+1}) \varepsilon'_t \right) \left(E \left(\varepsilon_{t+1} \varepsilon'_{t+1} \right) \right)^{-1} \quad (\text{E.1})$$

Conjecture that the proposition is correct, and thus a fixed point exists with $P = Q \Rightarrow M = 0$. Recall that since both conditions are identical in the parallel problem of Proposition 1, so we set $F_c = H_c = 0$. Since this fixed point replicates full information, we have, using (22), and (18)

$$\begin{aligned} \beta(Q) &= F_u S F'_u H'_\xi (H'_\xi F_u S F'_u H'_\xi)^{-1} \\ &= F_u S F'_u H'_\xi (F'_u H'_\xi)^{-1} S^{-1} (H'_\xi F_u)^{-1} \\ &= F_u (H'_\xi F_u)^{-1} \end{aligned} \quad (\text{E.2})$$

and hence

$$\beta(Q) H'_\xi Q = F_u (H'_\xi F_u)^{-1} H'_\xi F_u S F'_u = F_u S F'_u = Q \quad (\text{E.3})$$

$$\Rightarrow M(Q) = (I_r - \beta(Q) H'_\xi) Q = 0 \quad (\text{E.4})$$

$$\Rightarrow P = Q \quad (\text{E.5})$$

hence $n = s$ and $|H'_\xi F_u| \neq 0$ are necessary conditions for $P = Q$ to be a fixed point.

However, using our proof of Corollary 2, stability of a fixed point that replicates full information requires the third condition in Proposition 3, that the matrix $(I_r - \tilde{\beta}(Q) H'_\xi) F_\xi$ have eigenvalues with real parts strictly less than one in absolute value to ensure stability of the recursion (and hence, via Corollary 2, of the filtering errors, f_t). This condition is non-trivial because the same Corollary shows that this matrix may have up to $n - r$ non-zero eigenvalues. ■

E.2 Equivalence of Proposition 3 with Fernandez-Villaverde et al (2007)

Fernandez-Villaverde et al (2007) analyse the problem of econometric invertibility: ie, whether an econometrician can infer true structural shocks and impulse responses from an estimated vector autoregressive representation. It turns out that the conditions in which their “square case” is invertible correspond precisely to those given in Proposition 3. They do not, however, draw out the link with pre-determined

variables, nor with the stability of the Kalman filter in the neighbourhood of the fixed point corresponding to full information.

To show the correspondence, Fernandez-Villaverde et al have

$$x_{t+1} = Ax_t + Bw_{t+1} \quad (\text{E.6})$$

$$y_{t+1} = Cx_t + Dw_{t+1} \quad (\text{E.7})$$

where x_t is $n \times 1$, y_t is $k \times 1$ and w_t is $m \times 1$. Note that the timing convention in (E.7) is somewhat unusual compared to the standard Kalman filter state space representation, in which the measured variables are normally related to the contemporaneous states (possibly plus noise).⁴ They then assume

$$m = k \quad (\text{E.8})$$

$$|D| \neq 0 \quad (\text{E.9})$$

implying

$$w_{t+1} = D^{-1} [y_{t+1} - Cx_t] \quad (\text{E.10})$$

which gives, after substitution into (E.6),

$$x_{t+1} = [A - BD^{-1}C] x_t + BD^{-1}y_{t+1} \quad (\text{E.11})$$

which gives the third condition (their Condition 1, since the first two conditions are taken as implicit)

$$|\Lambda_i [A - BD^{-1}C]| < 1 \quad (\text{E.12})$$

where the Λ_i are eigenvalues. If satisfied, this means that x_{t+1} can be recovered from the infinite history of y_{t+1} (implicitly, even if $n > m = k$, which will be the case, as we note, if there are predetermined variables).

We have (using (6) and (20)), for the states,

$$\xi_{t+1} = F_\xi \xi_t + F_u u_{t+1} \quad (\text{E.13})$$

which with appropriate relabelling of variables and matrices is precisely the same as (E.6). Our measurement equation can be written as

$$y_{t+1} = H_\xi \xi_{t+1} \quad (\text{E.14})$$

⁴This is however the same timing convention as used by, eg, Whittle, 1983, p 146.

which, substituting from (E.13) gives

$$y_{t+1} = H'_\xi F_\xi \xi_t + H'_\xi F_u u_{t+1}$$

which, with appropriate relabelling (ie, $C = H'_\xi F_\xi$, $D = H'_\xi F_u$) is identical to (E.7). The relabelling makes it clear that their first two conditions are identical to ours. Our third condition relates to the eigenvalues of

$$[I - \beta(Q) H'_\xi] F_\xi \tag{E.15}$$

but in (67) of our appendix we show that

$$\beta(Q) = F_u (H'_\xi F_u)^{-1} \tag{E.16}$$

ie, can be derived directly from structural matrices. Substituting the crucial matrix becomes

$$F_\xi - F_u (H'_\xi F_u)^{-1} H'_\xi F_\xi \tag{E.17}$$

which given the relabelling can be written as

$$A - BD^{-1}C. \tag{E.18}$$

F Proof of Proposition 4

F.1 Proof of Part a)

We first exploit the diagonalisation $G = V\Lambda V^{-1}$ used in Assumption 2, which allows us to represent the process for the estimated states (19), conditional upon the Kalman Filter having converged, in the diagonalised autoregressive form

$$X_t \equiv V^{-1} \widehat{\xi}_t = \Lambda X_{t-1} + \Gamma \varepsilon_t \tag{F.1}$$

with $\Gamma = V^{-1}\beta$. Note that this subsumes all invertible cases in which we can write $\widehat{\xi}_t = \xi_t$; $\beta \varepsilon_t = F_u u_t$, hence applies irrespective of the nature of the information set. We assume in what follows, without loss of generality, that the diagonal elements of Λ , the eigenvalues of G , are non-zero and distinct (ie that we have a minimal state variable representation). We can then rewrite the measurement equation (8) as

$$i_t = \Psi' X_t \tag{F.2}$$

where, exploiting Corollary 4, $\Psi' = (H'_\xi + H_c\eta')V$. Note that we have, using (21) and (D.9),

$$(H'_\xi + H_c\eta')\beta = (H'_\xi + H_c\eta')\tilde{\beta} \left[I + H_c\eta'\tilde{\beta} \right]^{-1} = \left[I + H_c\eta'\tilde{\beta} \right] \left[I + H_c\eta'\tilde{\beta} \right]^{-1} = I$$

and hence

$$\Psi'\Gamma = I \tag{F.3}$$

Partition Ψ' and order the elements of X_t such that we can write

$$\Psi' = \begin{bmatrix} n \times n & n \times (r-n) \\ \Psi_1 & \Psi_2 \end{bmatrix}; \quad |\Psi_1| \neq 0 \Rightarrow X_{1t} = \Psi_1^{-1} [i_t - \Psi_2 X_{2t}]$$

hence, exploiting the property that Λ is diagonal,

$$\begin{aligned} i_t &= \Psi' \left\{ \begin{bmatrix} \Lambda_{11} & 0 \\ 0 & \Lambda_{22} \end{bmatrix} \begin{bmatrix} \Psi_1^{-1} [i_{t-1} - \Psi_2 X_{2t-1}] \\ X_{2t-1} \end{bmatrix} + \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \end{bmatrix} \varepsilon_t \right\} \\ &= \begin{bmatrix} \Psi_1 & \Psi_2 \end{bmatrix} \left\{ \begin{bmatrix} -\Lambda_{11}\Psi_1^{-1}\Psi_2 \\ \Lambda_{22} \end{bmatrix} [I - \Lambda_{22}L]^{-1} \Gamma_2 \varepsilon_{t-1} + \begin{bmatrix} \Lambda_{11}\Psi_1^{-1} \\ 0 \end{bmatrix} i_{t-1} \right\} + \varepsilon_t \\ &= \Phi_1 [I - \Lambda_{22}L]^{-1} \Gamma_2 \varepsilon_{t-1} + \Phi_2 i_{t-1} + \varepsilon_t \end{aligned} \tag{F.4}$$

where the last two lines use (F.3), and

$$\Phi_1 = \Psi' \begin{bmatrix} -\Lambda_{11}\Psi_1^{-1}\Psi_2 \\ \Lambda_{22} \end{bmatrix}; \quad \Phi_2 = \Psi_1 \Lambda_{11} \Psi_1^{-1} \tag{F.5}$$

Hence we can write

$$i_t = \frac{\Phi_1 \Lambda_{22}^*(L)}{\det [I - \Lambda_{22}L]} \Gamma_2 \varepsilon_{t-1} + \Phi_2 i_{t-1} + \varepsilon_t \tag{F.6}$$

where $\Lambda_{22}^*(L) = \text{adj} [I - \Lambda_{22}L]$. Note that $\det [I - \Lambda_{22}L]$ is a scalar polynomial of order $r - n$, and the elements of the (diagonal) matrix polynomial $\Lambda_{22}^*(L)$ are scalar polynomials of order $r - n - 1$.⁵ Hence

$$\lambda(L) i_t = \theta(L) \varepsilon_t \tag{F.7}$$

⁵Letting $\{\mu_i\}_{i=1}^{r-n}$ be the eigenvalues of Λ_{22} , then $\det [I - \Lambda_{22}L] = \prod_{i=1}^{r-n} (1 - \mu_i L)$, and the i th diagonal element of $[I - \Lambda_{22}L]^{-1} = (1 - \mu_i L)^{-1}$, hence the i th diagonal element of $\Lambda_{22}^*(L)$ equals $\prod_{j \neq i} (1 - \mu_j L)$, a polynomial of order $r - n - 1$.

where

$$\lambda(L) = \det [I - \Lambda_{22}L] (I_n - \Phi_2L) \quad (\text{F.8})$$

$$\theta(L) = \det [I - \Lambda_{22}L] I_n + \Phi_1 \Lambda_{22}^*(L) \Gamma_2 L \quad (\text{F.9})$$

which is a VARMA(n, p, q) representation with $p = r - n + 1$ and $q = r - n$ with both holding with equality because, as assumed at the outset, the diagonal elements of Λ (the eigenvalues of G) are real and distinct. If this is not the case, but G has $\tilde{r} < r$ distinct non-zero eigenvalues, the argument goes through replacing r with \tilde{r} throughout. The values of p and q given in the Proposition are the maximal orders of the minimal VARMA representation, in which the both the autoregressive and moving average matrix lag polynomials are in general non-diagonal. If these polynomials are restricted to be diagonal (ie, the system is a set of univariate ARMA representations with correlated innovations) this is equivalent to setting $n = 1$. ■

F.2 Proof of Part b)

In the case of instantaneous invertibility we can derive the states using (12), then substitute into (7) giving a vector autoregression for i_t . with innovations given by $\varepsilon_t = [H'_\xi + H_c \eta'] F_u u_t$, where ε_t and u_t are both of dimension r . In the case of asymptotic invertibility we have a generalised version of (23), given $H_c \neq 0$, of the form

$$\varepsilon_{t+1} = H'_\xi (\xi_{t+1} - E_t \xi_{t+1}) + H_c (c_{t+1} - E_t c_{t+1}) = [H'_\xi + H_c \eta'] F_u u_t$$

where the only difference is that ε_t and u_t are in this case both of dimension $n < r$. ■

F.3 Proof of Parts c), d) and e).

We first derive a (non-invertible) VARMA representation in terms of the vector of underlying structural innovations u_t by exploiting the structure of the proof of Part a) of the Proposition, with a relatively small number of changes.

Redefine X_t in (F.1) to be the minimal state vector for the true process in (25), which is driven by the true structural innovations u_t . Thus we now write

$$X_t = \Lambda X_{t-1} + \Gamma u_t \quad (\text{F.10})$$

From Corollary 2 the autoregressive matrix $(I - \tilde{\beta} H'_\xi) F_\xi$ of the filtering error process f_t has at most $r - n$ non-zero eigenvalues. Hence for $F_c \neq 0$, X_t will be of

dimension $2r - n$, and we can redefine V and Λ such that

$$\begin{bmatrix} \xi_t \\ f_t \end{bmatrix} = VX_t; \quad (\text{F.11})$$

$$V\Lambda V^{-1} = \begin{bmatrix} G & -F_c\eta' \\ 0 & (I - \tilde{\beta}H'_\xi)F_\xi \end{bmatrix} \quad (\text{F.12})$$

where V is now $2r \times (2r - n)$ and hence V^{-1} is a generalised inverse. The diagonal elements of Λ (which become the reciprocal autoregressive roots of the VARMA representation) are given by the non-zero eigenvalues of G and of $(I - \tilde{\beta}H'_\xi)F_\xi$. We redefine Γ consistently with (25) such that

$$\Gamma = V^{-1} \begin{bmatrix} I \\ (I - \tilde{\beta}H'_\xi) \end{bmatrix} F_u \quad (\text{F.13})$$

We also need to redefine Ψ' in (F.2), using (13) to write $c_t = \eta'(\xi_t - f_t)$, giving

$$\Psi' = \begin{bmatrix} H'_\xi + H_c\eta' & -H_c\eta' \end{bmatrix} V \quad (\text{F.14})$$

As a result (F.3) no longer holds, but becomes

$$\Psi'\Gamma = \begin{bmatrix} H'_\xi + H_c\eta' & -H_c\eta' \end{bmatrix} \begin{bmatrix} I \\ (I - \tilde{\beta}H'_\xi) \end{bmatrix} F_u \quad (\text{F.15})$$

reflecting the impact of structural shocks on both states and filtering errors.

Note that in the restricted case that the states are dynamically exogenous ($F_c = 0$) by inspection of (F.12) the true states are block recursive (there is no "contamination" of state dynamics by filtering errors), hence the minimal state vector of the true process remains of dimension r , as for the observable process.

Given the redefinitions above we can proceed as in the proof of Part a), which allows us to derive a representation of the form⁶

$$\lambda(L) i_t = [\det [I - \Lambda_{22}L] \Gamma + \Phi_1 \Lambda_{22}^*(L) \Gamma_2 L] u_t \quad (\text{F.16})$$

where now for the general case with $F_c \neq 0$, the autoregressive polynomial $\lambda(L)$ is of order $2(r - n) + 1$. The right-hand side defines a vector moving average process of order $2(r - n)$, of dimension n , in terms of an underlying vector of innovations u_t of dimension $s \geq n$

⁶Which implicitly defines $A(L)$ in the exposition in the main text.

The proof of Part c) of the proposition is now straightforward. A non-fundamental VARMA representation takes the general form

$$\lambda(L) i_t = \theta^{nf}(L) \eta_t \quad (\text{F.17})$$

where η_t is an n -vector of innovations that cannot be recovered from the history of i_t (ie, $\theta^{nf}(L)$ is non-invertible in non-negative powers of L), and $\lambda(L)$ and $\theta^{nf}(L)$ are both finite order polynomials. For the case where $n = s = \text{rank}(H'F_u)$, it follows immediately, by comparison of (F.16) and (F.17), that u_t is (up to a scaling factor) the vector of innovations to a non-fundamental VARMA representation. ■

By substitution from (F.17) into (F.16) we have, for any nonfundamental representation of the same moving average order⁷

$$\theta^{nf}(L) \eta_t = [\det [I - \Lambda_{22}L] \Gamma + \Phi_1 \Lambda_{22}^*(L) \Gamma_2 L] u_t \quad (\text{F.18})$$

For the case in Part c) where $n = s$, this simply defines a particular η_t and $\theta^{nf}(L)$. For the general case with $n < s$, inversion of $\theta^{nf}(L)$ involves terms in powers of $F = L^{-1}$, the forward shift operator. Hence non-fundamental innovations can only be recovered from an information set consisting of both the history and future of u_t , thus proving part d) of the Proposition. ■

Following Lippi & Reichlin's (1994) terminology, if the representation is of higher autoregressive moving average order than the fundamental representation it is a "nonbasic" representation. By comparison of the order of the representations in (F.7) and (F.16) it is evident that for the general case (ie, with $F_c \neq 0$) the true process results in a VARMA($n, 2(r - n) + 1, 2(r - n)$) representation, hence there is an associated set of non-fundamental representations of the same order, defined by (F.18). By comparison the fundamental representation in (F.7) is VARMA($n, r - n + 1, r - n$); hence the non-fundamental representations defined by (F.18) are also non-basic, proving Part e) of the Proposition. ■

⁷Note that even if we fix the moving average order there are still in general multiple (but a finite number of) non-fundamental representations, that can be generated by sequentially replacing a single MA root with its reciprocal, but in which at least one such root is less than unity in absolute value (the representation with all roots greater than unity in absolute value is the unique fundamental representation).