DYNAMIC HIGHER ORDER EXPECTATIONS

KRISTOFFER P. NIMARK

Abstract. In models where privately informed agents interact, they may need to form higher-order expectations, i.e. expectations about other agents' expectations. In this paper we prove that there exists a unique equilibrium in a class of linear dynamic rational expectations models in which privately informed agents form higher order expectations. We propose an iterative procedure that recursively computes increasing orders of expectations. The algorithm is a contraction mapping, and the implied dynamics of the endogenous variables converge to the unique equilibrium of the model. The contractive property of the algorithm implies that, in spite of the fact that the model features an *infinite regress of expectations*, the equilibrium dynamics of the model can be approximated to an arbitrary accuracy with a finite-dimensional state. We provide explicit bounds on the approximation errors.

Keywords: Dynamic Higher Order Expectations, Private Information, Asset Pricing, Infinite Regress of Expectations

1. Introduction

Many economic decisions involve predicting the actions of other agents. For instance, firms in oligopolistic markets may need to predict how much their competitors are investing in productive capacity and traders in financial markets may need to predict how much other traders will pay for an asset in the future. In settings where all agents are identical and share the same information, this becomes a trivial problem: An individual agent can predict the behavior of other agents by introspection, since all agents will choose the same action in equilibrium. The problem becomes more difficult, but also more realistic, if the common information assumption is relaxed. Predicting the actions of others is then distinct from predicting ones own actions. But since other agents face a symmetric problem, each individual agent needs to predict other agents' expectations about the actions of other agents, and so on, leading to the well-known infinite regress of expectations problem, e.g. Townsend (1983).

As first pointed out by Townsend (1983) and later emphasized by Sargent (1991), natural state representations of dynamic models with privately informed agents tend to become infinite-dimensional. As a consequence, little is known about the equilibrium properties of

Date: First version November 2006, substantially revised on March 2, 2017. The author thanks Francisco Barillas, Vasco Carvalho, David Easley, Jesus Fernandez-Villaverde, Chris Huckfeldt, Alex Kohlhas, Gabor Lugosi, Filip Matejka, Christian Matthes, Stefan Pitschner, Mirko Wiederholt, Thomas Sargent, Iian Smyth and seminar participants at New York University, Birkbeck College, Goethe University Frankfurt, the 2007 annual meeting of the Society for Economic Dynamics in Prague, Australian Workshop for Macro Dynamics, Institute for International Economic Studies at Stockholm University, University of Amsterdam, and the 2009 Econometric Society meeting in San Francisco for useful comments and suggestions.

Address: Economics Department, Cornell University, Ithaca 14853 NY. *e-mail*: pkn8@cornell.edu *web page*: www.kris-nimark.net .

this type of model and no general solution method has been developed. This paper makes two contributions towards filling these gaps. First, we prove that there exists a unique equilibrium in a class of linear rational expectations models with privately informed agents. Second, we propose an algorithm that can be used to approximate the equilibrium of models that feature an infinite regress of expectations to an arbitrary accuracy with a finite-dimensional state and explicit approximation error bounds. These results hold under quite general conditions: It is sufficient that agents discount the future and that the exogenous processes follow stationary (but otherwise unrestricted) VARMA processes.

The strategy we take is the following. To prove that there exists a unique equilibrium we show that the mapping defined by the filtering and prediction problem of the agents is a contraction on the space relevant for the endogenous processes. We then propose a procedure that iterates on the model's Euler equation, while recursively increasing the number of orders of expectations that are taken into account. The proposed procedure is of the form that is shown to be a contraction, and can thus be used to find the equilibrium dynamics of the model.

To understand how the results described above help us overcome the infinite regress of expectations problem, it is worthwhile describing the iterative algorithm in some detail. As a starting point, consider agents that engage in only first order reasoning. That is, they do not consider the effect of other agents' expectations on the equilibrium outcome when they solve their filtering and prediction problem. Substituting in the expectations of the naive agents into the equilibrium conditions of the model results in a process for the endogenous variable that depends on agents' first order expectations about the exogenous processes. In the second step of the algorithm, agents take the endogenous process implied by the first step as given when they solve their filtering and prediction problem. Because that process depends on other agents' expectations, agents then need to engage in second order reasoning and form expectations about other agents' expectations. At the k-step of the iteration, agents take the endogenous dynamics implied by $k-1$ order reasoning as given when they solve their filtering and prediction problem. To form higher order expectations, it is necessary that agents are endowed with a theory of how other agents form expectations. Here, we assume that agents form model consistent (i.e. rational) expectations and that this is common knowledge. This gives enough structure to compute the dynamics of the endogenous variables, while taking into account expectations of an arbitrarily high order.[1]

By itself, this procedure does not solve the infinite regress of expectations problem. However, each step in the iterative procedure is associated with an implied process for the endogenous variables. Because the iterative procedure is a contraction mapping on the space relevant for the endogenous processes, these converge to the unique equilibrium of the model as the number of iterations increase. The contractive property of the algorithm also implies that the equilibrium dynamics of the model can be approximated to an arbitrary accuracy

[1]In the terminology established by Harsanyi (1967-8), common knowledge of rational expectations here means that there is a common prior about the true state of nature and the joint probability distribution of the true state of nature and the "types". Different "types" are distinguishable only by the realizations of the private signals that they have observed in the past. The common prior then endows agents with sufficient knowledge to form model consistent expectations of the signals observed by other agents.

with a finite-dimensional state and we provide explicit approximation error bounds as a function of the maximum order of expectation considered.

Finite numbers can still be very large, and one may ask if these results are relevant in practice. Using a simple asset pricing model we demonstrate numerically that the equilibrium dynamics can be captured by a low number of orders of expectations, i.e. by a vector of dimension in the single digits. This latter result may be reassuring to those who on grounds of human cognitive constraints doubt that economic agents form an infinite hierarchy of higher-order expectations, e.g. Nagel (1995).

Introducing information imperfections into dynamic macroeconomic and finance models is not a new idea. Well-known early references include Phelps (1970), Lucas (1972, 1973, 1975), Townsend (1983), Singleton (1987) and Sargent (1991). However, recently, there has been a renewed interest in the topic and several interesting results have emerged. First, private information about quantities of common interest to all agents have been shown to introduce inertia and sluggishness of endogenous variables in settings with strategic complementarities, e.g. Woodford (2002), Morris and Shin (2006), Nimark (2008), Mackowiak and Wiederholt (2009) and Angeletos and La'O (2009). Second, noisy public signals can provide a plausible theory of observed co-movement of aggregate variables as in Lorenzoni (2009). Third, private information may also have normative policy implications as shown by Angeletos and Pavan (2007), Lorenzoni (2010) and Paciello and Wiederholt (2013). Fourth, in financial markets, private information may introduce speculative behavior akin to the "beauty contest" metaphor of Keynes (1936), e.g. Allen, Morris and Shin (2006), Bacchetta and van Wincoop (2006), Grisse (2009), Cespa and Vives (2012), Kasa, Walker and Whiteman (2014), Rondina and Walker (2014) and Barillas and Nimark (2016a, 2016b). In spite of the renewed interest, no general solution methodology with known properties has emerged for solving this class of models. In the absence of a general solution method, existing approaches have imposed additional restrictions to simplify the prediction problem faced by agents. To understand the contribution of the present paper it is worthwhile to spell out these assumptions explicitly.

The first approach used to solve dynamic models with private information exploited the idea that if private information is short-lived, it is often possible to avoid modeling higher-order expectations explicitly as state variables. Lucas (1972,1975) achieved this by assuming that agents pool their information between periods. More recently, Allen, Morris and Shin (2006), Cespa and Vives (2012) or Banerjee, Kaniel and Kremer (2009) analyze finite horizon models in which agents use non-recursive methods to predict endogenous variables directly by conditioning on a finite history of signals. Another way to make private information short-lived is to assume that all shocks are observed perfectly by all agents with a lag. This assumption was first introduced by Townsend (1983) as a way to restrict the dimension of the relevant state for 'forecasting the forecasts of others'. Optimal forecast of any variable of interest can then be constructed using projections onto the perfectly revealed state and a finite-dimensional vector of private signals. Variations of this method has also been used by Singleton (1987), Hellwig (2002), Lorenzoni (2009) and Hellwig and Venkateswaran (2009).

In models where agents do not face any intertemporal optimality conditions, it is often possible to simplify the state. For instance, Woodford (2002) solves a model in which agents face a dynamic filtering problem, but where the decision problem is static. The endogenous variable in Woodford's model is a function of a geometric average of higher-order expectations about the current fundamental. In this set-up, Woodford is able to derive an exact solution to his model as a function of a two-dimensional state. Huo and Takayama (2015a) propose an alternative way to solve a model of the same static decision form as Woodford's. Their method is exact when agents observe only exogenous signals with finite order ARMA representations.[2]

Kasa, Walker and Whiteman (2014) solve a model that, like the class of models analyzed in the present paper, features a dynamic equilibrium condition in the form of an Euler equation. They solve directly for the MA representation of the equilibrium dynamics of their model by imposing that the equilibrium representation is non-invertible. This method is further developed in Rondina and Walker (2014). Solving directly for the equilibrium as a function of the white noise innovations of the model is elegant, but requires restrictive assumptions on the ARMA structure of the observed variables. To solve their model, Rondina and Walker require the endogenous variable to follow a low order ARMA process that admits both an invertible and non-invertible representation.[3] Huo and Takayama (2015b) solve a dynamic RBC model with capital accumulation, but their solution method is not applicable to models in which agents observe signals that are functions of endogenous variables.

In contrast to existing approaches, the method proposed here allows for the exogenous processes to follow any stationary VARMA process and agents that observe both endogenous and exogenous signals. It is also applicable to models with multiple endogenous variables. The approach has several advantages. First, since fewer modeling compromises are needed, the method allows us to solve empirically more plausible models that more closely resembles the full information models in the quantitative macroeconomics and finance literature. Second, the solution method is both flexible enough and computationally fast enough to use for likelihood based estimation as demonstrated by Melosi (2016) and Nimark (2014), who solve and estimate business cycles models with privately informed agents, and by the empirical asset pricing papers by Barillas and Nimark (2016a, 2016b) and Struby (2016). The method thus makes it feasible to empirically quantify the importance of private information and provides a bridge between the large theoretical literature and the data. Third, modeling higher-order expectations as state variables helps intuition and makes the link between private information and the dynamics of the endogenous variables transparent. Fourth, the method relies on standard tools used in macroeconomics and time series, requiring no more than basic proficiency with the Kalman filter.

The next section presents an asset pricing model with privately informed agents. While simple, the model serves as an archetype to illustrate the complications that arise from the infinite regress of expectations and how these can be overcome by the proposed method. Section 3 presents an algorithm that iterates on the model's Euler equation while recursively increasing the number of orders of expectations considered. In Section 4 we prove that there

[2]To solve their static decision model with endogenous signals, Huo and Takayama (2015a) rely on an ad hoc approximation method with unknown properties.

[3]To date, Rondina and Walker's (2014) method has only been applied to models featuring a single endogenous variable and where an agent's private signal is a sum of two transitory shocks rather than a noisy measure of a persistent process. However, it is not clear that these are fundamental restrictions implied by their approach.

exists a unique solution to the model. There, we also show that the proposed algorithm is a contraction that converges to the unique solution of the model, and that the equilibrium dynamics can be approximated to an arbitrary accuracy by a finite-dimensional state. Section 5 illustrates how the solution method can be used in practice by solving the simple asset pricing model. Section 6 contains the main theorem of the paper that extends the results of Section 4 to a more general class of models. Section 7 concludes.

2. An Archetypal Dynamic Model

In this section we present a simple asset pricing model populated with privately informed agents. It is arguably the simplest dynamic set-up in which the infinite regress of expectations arises, and we will use it to illustrate the proposed solution method. The results derived using the simple model are extended to a more general class of models in Section 6.

2.1. **A simple dynamic model structure.** The model is a simplified version of Singleton's (1987) CARA utility overlapping generations asset pricing model. As in Singleton (1987), agents observe a private signal that is useful for predicting future asset prices. However, we relax Singleton's assumption that all shocks are revealed perfectly with a two period lag and instead assume that the shocks are never revealed. Time is discrete and indexed by t and there is a continuum of agents indexed by $j \in (0,1)$. The equilibrium price of the asset is determined by an Euler equation of the form

$$p_t = \beta \int E\left[p_{t+1} \mid \Omega_{t,j}\right] \, dj - (\theta_t + \varepsilon_t) : \varepsilon_t \sim N\left(0, \sigma_\epsilon^2\right) \tag{2.1}$$

where $\Omega_{t,j}$ is the information set of agent j and $0 \leq \beta < 1$. Asset supply is stochastic and given by the sum of the transitory shock ε_t and the persistent component θ_t that follows

$$\theta_t = \rho\theta_{t-1} + u_t : u_t \sim N\left(0, \sigma_u^2\right) \tag{2.2}$$

where $0 \leq |\rho| < 1$. Agent j observes the signal vector $\mathbf{s}_{t,j}$

$$\mathbf{s}_{t,j} \equiv \begin{bmatrix} z_{t,j} \\ p_t \end{bmatrix} \tag{2.3}$$

where $z_{t,j}$ is a noisy private signal about the exogenous fundamental θ_t of the form

$$z_{t,j} = \theta_t + \eta_{t,j} : \eta_{t,j} \sim N\left(0, \sigma_\eta^2\right). \tag{2.4}$$

Solving the model requires expressing the average expectations term $\int E\left[p_{t+1} \mid \Omega_{t,j}\right] dj$ in the Euler equation (2.1) as a function of the exogenous processes. When forming expectations, agents' condition on the history of both the price p_t and the private signal $z_{t,j}$. Agent j's information set is thus defined by the filtration $\Omega_{t,j} \in \left\{z_{t,j}, p_t, \Omega_{t-1,j}\right\}$. Because p_t depends on both θ_t and ε_t, agents will in general not be able to back out θ_t and ε_t perfectly from observing p_t. The private signal $z_{t,j}$ is then useful for predicting the next period price p_{t+1}. To understand the implications of introducing private information in this setting, it is useful to first show how the model can be solved under full information.

2.2. **The full information solution.** If all agents observe θ_t directly so that $\theta_t \in \Omega_t^j$ for all j, the Euler equation (2.1) can be used to recursively substitute out expectations about p_{t+1}, $p_{t+2,...}$ The price can then be expressed as a discounted sum of expected future values of θ_t

$$p_t = \sum_{k=0}^{\infty} \beta^k E\left(\theta_{t+k} \mid \theta_t\right) + \varepsilon_t. \tag{2.5}$$

For any random variable X, the law of iterated expectations states that

$$E\left(E\left[X \mid \Omega'\right] \mid \Omega\right) = E\left(X \mid \Omega\right) \tag{2.6}$$

if and only if $\Omega \subseteq \Omega'$. Under full information, we can use that the common information set Ω_t is defined by the filtration

$$\Omega_t = \{\theta_t, \Omega_{t-1}\} \tag{2.7}$$

so that $\Omega_t \subseteq \Omega_{t+k}$ for all $k \geq 0$. The law of iterated expectations (2.6) together with (2.2) thus implies that

$$\theta_{t+k} = \rho^k \theta_t. \tag{2.8}$$

The infinite sum (2.5) can then be simplified to

$$p_t = \frac{1}{1 - \beta\rho}\theta_t + \varepsilon_t \tag{2.9}$$

if $|\beta\rho| < 1$.

2.3. **Private information and a complication.** We want to solve the model above without imposing that all agents have access to the same information. With privately informed agents, we can still substitute (2.1) forward. After the first step we get

$$p_t = \varepsilon_t + \theta_t + \beta \int E\left[\theta_{t+1} \mid \Omega_t^j\right] dj + \beta^2 \int E\left[\int E\left[p_{t+2} \mid \Omega_{t+1}^{j'}\right]\ dj' \mid \Omega_t^j\right]\ dj. \tag{2.10}$$

The price now depends on the exogenous supply shocks ε_t and θ_t, the average expectation in period t of θ_{t+1} but also on higher-order expectations. The last term on the right hand side of (2.10) is the average expectation in period t of the average expectation in period $t+1$ of the price in period $t+2$. Continued recursive substitution of (2.1) gives the price as a discounted sum of higher-order expectations about future values of θ_t

$$\begin{aligned} p_t \;\; = \;\; & \varepsilon_t + \theta_t + \beta \int E\left[\theta_{t+1} \mid \Omega_t^j\right] dj + \beta^2 \int E\left[\int E\left[\theta_{t+2} \mid \Omega_{t+1}^{j'}\right]\ dj' \mid \Omega_t^j\right]\ dj + ... \\ & ... + \beta^k \int E\left[\cdots \int E\left[\theta_{t+k} \mid \Omega_{t+k-1}^{j''}\right]\ dj'' \cdots \mid \Omega_t^j\right]\ dj + ... : k \to \infty \end{aligned} \tag{2.11}$$

That is, p_t depends on the average expectation in period t of θ_{t+1}, the average expectation in period t of the average expectation in period $t+1$ of θ_{t+2}, and the average expectation in period t of the average expectation in period $t+1$, and so on, of the average expectations in period $t+k-1$ of θ_{t+k}, with k tending to infinity. It is the fact that the price depends on these higher-order expectations that makes standard solution methods inapplicable.

As has been noted before, e.g. Allen, Morris and Shin (2006), higher-order expectations generally differ from average first order expectations. Formally, it is because the current information set of an agent is not nested in the future information sets of other agents that makes the law of iterated expectations (2.6) inapplicable to compute the higher-order expectation in (2.11). To see why this matters, note that the law of iterated expectations can be interpreted as a consequence of an agent's inability to predict their own forecast errors. Their expectation about their future expectations about θ_{t+k} must then coincide with their current expectations about θ_{t+k}.

The same is not true about expectations about other agents' expectations. Because individual agents observe private signals, they can systematically predict the future forecasting errors of other agents. By definition, an individual agents' prediction about other agents' forecasting error is simply the difference between his first and second order expectation. That individual agents can predict other agents' forecasting errors is thus equivalent to the statement that first and second order expectations do not coincide. A similar argument can be applied to higher-order expectations. Because different orders of expectations about future values of θ_t do not generally coincide, we cannot reduce the discounted sum in (2.11) to a function of a single period t state variable.

2.4. **Signal extraction from endogenous variables with private information.** An alternative to computing the higher-order expectations about θ_{t+k} in (2.11) is to let agents directly predict the next period price and then take the average across agents. If the individual agents' expectations are rational given their information sets, this approach will yield a solution to the model. A "brute force" approach to this problem would be to project the next period price onto the entire history of signals observed by the agents. However, this would imply conditioning on a vector of signals with an ever-increasing dimension as time passes, which is impractical.

As described by both Townsend (1983) and Sargent (1991), a natural alternative is to write down the prediction problem recursively, letting agents update only to the new information that arrives in period t using the Kalman filter. However, an agent that wants to use the information contained in an endogenous variable to extract information about a latent state variable of common interest will need to include an infinite number of higher-order expectations as state variables. The logic of their argument in the context of the model (2.1)-(2.4) is as follows.

Consider an agent that wants to form an expectation about the next periods price p_{t+1}. He clearly needs to form an expectation about θ_{t+1} which by (2.2) depends in the current value of θ_t. The agent can observe the current price p_t, and because the price depends on θ_t, the agent would like to extract information about θ_t from p_t. But since the price also depends on the average expectations of other agents, who are solving the same filtering and prediction problem, the price p_t will depend not only on θ_t but also on the average expectation about θ_t. Our agent thus needs to include the average expectation about θ_t in his state to effectively extract the information from p_t. But again, since the agent knows that all other agents face the exact same problem, the price will then depend on the average expectation of the average expectation about θ_t, and the agent then needs to include also the average expectation of the average expectation in his state to effectively extract the information from p_t. This recursion never ends, and natural recursive state representations in this kind of set-up thus tend to be infinite-dimensional as higher and higher orders of expectations need to be considered. It was this fact that made Townsend (1983) look for representation that did not explicitly include higher-order expectations. Instead, Townsend assumed that the latent state is revealed perfectly with a lag. We would like to be able to solve the this type of model without assuming that the state is ever perfectly revealed.

2.5. **The strategy.** Our strategy to overcome the infinite regress of expectations problem is as follows. First, we present a recursive algorithm that iterates on the Euler equation (2.1). The algorithm follows the logic described in the previous paragraph closely and incrementally adds one order of expectation at each step in the iteration. If the paper ended there, not much would have been achieved, as the dimension of the state would grow arbitrarily large. However, we then prove that there exists a unique equilibrium of the model described by (2.1)-(2.4) and that the algorithm presented in the next section converges to this equilibrium. The contractive property of the algorithm also implies that the equilibrium dynamics can be approximated arbitrarily well with a finite-dimensional state.

3. Recursively computing higher-order expectations

This section presents an algorithm that iterates on the Euler equation (2.1) while incrementally increasing the number of orders of expectations included in the model's representation. One way to think of the algorithm is as starting from the filtering and prediction problem of a "naive" agent who, in the language of Nagel (1995), only engage in "first order reasoning". That is, the naive agent solves his filtering and prediction problem without taking into account other agents' expectations. Substituting the cross-sectional average expectation of similarly naive agents into the Euler equation (2.1) yields a price function that depends only on the exogenous variables and average first order expectations. In the next step, we let agents take this price function as given and thus take into account that other agents' first order expectations affect the price that they observe. In the second step, agents then engage in second order reasoning when they solve their filtering and prediction problem. Extending this argument, in the k^{th} iteration agents engage in k order reasoning. The algorithm thus closely follows the logic used by Townsend (1983) and Sargent (1991) to illustrate the intractability of the infinite regress of expectations. Here we use the same logic to compute the price process taking into account an arbitrarily large number of orders of expectations.

3.1. **Notation and definitions.** We denote agent j's first order expectation of θ_t conditional on his period t information set $\Omega_{t,j}$ as

$$\theta_{t,j}^{(1)} \equiv E\left[\theta_t \mid \Omega_{t,j}\right]. \tag{3.1}$$

The average first order expectation $\theta_t^{(1)}$ is obtained by taking averages of (3.1) across agents so that

$$\theta_t^{(1)} \equiv \int E\left[\theta_t \mid \Omega_{t,j}\right] dj. \tag{3.2}$$

The k^{th} order expectation of θ_t is defined recursively as

$$\theta_t^{(k)} \equiv \int E\left[\theta_t^{(k-1)} \mid \Omega_{t,j}\right] dj \tag{3.3}$$

where we used the convention that the zero order expectation of θ_t is the actual value of the variable so that

$$\theta_t^{(0)} \equiv \theta_t. \tag{3.4}$$

Full information rational expectations implies that the variable θ_t is common knowledge so that $\theta_t = \theta_t^{(k)} : k = 1, 2, ...$ for all t. We refer to a sequence of expectations, for instance from order zero to k, as a *hierarchy* of expectations from order zero to k. Vectors consisting of a hierarchy of expectations are denoted $\theta_t^{(0:k)}$ and defined as

$$\theta_t^{(0:k)} \equiv \begin{bmatrix} \theta_t^{(0)} & \theta_t^{(1)} & ... & \theta_t^{(k)} \end{bmatrix}'. \tag{3.5}$$

3.2. **Filtering and prediction with first order reasoning.** To solve the model, we need to find the equilibrium cross-sectional average prediction of p_{t+1} in (2.1). As a starting point, consider the prediction problem of an agent that understands that the price is affected by $\theta_t + \varepsilon_t$, but is naive in the sense that he does not understand that the current price is also affected by the expectations of others. The agent thus believes that the price follows

$$p_t^0 = -\theta_t - \varepsilon_t \tag{3.6}$$

where the zero superscript indicates that the prices process only depends on the actual, i.e. zero order expectations of θ_t. Since θ_t follows an autoregressive process, in order to predict p_{t+1}^0, the agent thus needs to form an expectation about θ_t. Given the linear structure and Gaussian shocks, agent j's optimal expectation of $\theta_{t,j}$ conditional on the history of observed signals $\mathbf{s}_{t,j}$ is given by the Kalman filter update equation

$$\theta_{t,j}^{(1)} = \rho\theta_{t-1,j}^{(1)} + K_0 \left[\mathbf{s}_{t,j} - L_0\rho\theta_{t-1,j}^{(1)}\right] \tag{3.7}$$

associated with the state space system

$$\theta_t = \rho\theta_{t-1} + u_t \tag{3.8}$$

$$\mathbf{s}_{t,j} = D_0\theta_t + e_1\eta_{t,j} + e_2\varepsilon_t \tag{3.9}$$

where $D_0 = \begin{bmatrix} 1 & -1 \end{bmatrix}'$, K_0 is the Kalman gain and e_n is a conformable vector with a 1 in the n^{th} position and zeros elsewhere.

3.2.1. *Finding the average first order expectation.* To find the average expectation, use that $\int \eta_{t,j}dj = 0$ to integrate the signal vector (3.9) in (3.7) across agents. Combine the resulting expression with the true law of motion for θ_t and rearrange to get the law of motion for the hierarchy of expectation from order 0 to 1 as

$$\begin{bmatrix} \theta_t \\ \theta_t^{(1)} \end{bmatrix} = \begin{bmatrix} \rho & 0 \\ K_0D_0\rho & (I - K_0D_0)\,\rho \end{bmatrix} \begin{bmatrix} \theta_{t-1} \\ \theta_{t-1}^{(1)} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ K_0D_0 & K_0 \end{bmatrix} \begin{bmatrix} u_t \\ \varepsilon_t \end{bmatrix} \tag{3.10}$$

or more compactly

$$\theta_t^{(0:1)} = M_1\theta_{t-1}^{(0:1)} + N_1\mathbf{w}_t : \mathbf{w}_t \sim N(0, I). \tag{3.11}$$

Given the price process p_t^0, the average expectation of the price in the next period can then be computed as

$$\int E\left[p_{t+1}^0 \mid \Omega_{t,j}^0\right] \, dj = -\int E\left[\theta_{t+1} \mid \Omega_{t,j}^0\right] \, dj \tag{3.12}$$

$$= -\rho\theta_t^{(1)} \tag{3.13}$$

where the superscript on $\Omega_{t,j}^0$ denotes that $\Omega_{t,j}^0 = \left\{p_t^0, z_{t,j}, \Omega_{t-1,j}^0\right\}$.

3.2.2. *The price process with first order reasoning.* To find the price implied by first order reasoning, substitute the average expectation (3.13) into the Euler equation (2.1) to get the new price process p_t^1 as

$$p_t^1 = \mathbf{g}_1'\theta_t^{(0:1)} - \varepsilon_t \tag{3.14}$$

where

$$\mathbf{g}_1' = -\left[\begin{array}{cc} 1 & \beta\rho \end{array}\right]. \tag{3.15}$$

The law of motion (3.10) for $\theta_t^{(0:1)}$ together with the new measurement equation

$$\mathbf{s}_{t,j} = D_1\theta_t^{(0:1)} + e_1\eta_{t,j} + e_2\varepsilon_t \tag{3.16}$$

where

$$D_1 = \left[\begin{array}{c} e_1' \\ \mathbf{g}_1' \end{array}\right] \tag{3.17}$$

then defines a new state space system.

We can now see the beginning of the infinite regress problem described by Townsend (1983) and Sargent (1991). Because the average expectation $\theta_t^{(1)}$ is now part of the state that determines the price p_t^1, in order to use the information in the signal vector (3.16), agent j need to form an expectation about the average expectation of other agents.

3.3. **Recursively finding higher-order expectations.** Above, we derived the average expectation of agents that engage in first order reasoning taking the exogenous processes as given. The same steps can be used to describe the filtering and prediction problem faced by an agent that engages in $k+1$ order reasoning, taking the price process implied by k order reasoning as given. This strategy allows us to recursively compute the price process and the law of motion of the expectations hierarchy for arbitrarily high orders of expectations.

3.3.1. *The filtering problem.* Consider the problem of estimating the state $\theta_t^{(0:k)}$ conditional on the history of the price p_t^k

$$p_t^k = \mathbf{g}_k'\theta_t^{(0:k)} - \varepsilon_t \tag{3.18}$$

and the private signal $z_{t,j}$. Agent j's optimal estimate of the hierarchy $\theta_t^{(0:k)}$, that is, agent j's expectations from order 1 to $k+1$, is given by the Kalman update equation

$$\theta_{t,j}^{(1:k+1)} = M_k\theta_{t-1,j}^{(1:k+1)} + K_k\left[\mathbf{s}_{t,j} - D_kM_k\theta_{t-1,j}^{(1:k+1)}\right] \tag{3.19}$$

where K_k is the Kalman gain associated with the state space system

$$\theta_t^{(0:k)} = M_k\theta_{t-1}^{(0:k)} + N_k\mathbf{w}_t : \mathbf{w}_t \sim N(0, I) \tag{3.20}$$

$$\mathbf{s}_{t,j} = D_k\theta_t^{(0:k)} + R_w\mathbf{w}_t + R_\eta w_{t,j} : w_{t,j} \sim N(0,1). \tag{3.21}$$

The matrices D_k, R_w and R_η in the measurement equation (3.21) are defined as

$$D_k = \begin{bmatrix} e_1' \\ \mathbf{g}_k' \end{bmatrix}, \quad R_w = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_\varepsilon \end{bmatrix}, \quad R_\eta = \begin{bmatrix} \sigma_\eta \\ 0 \end{bmatrix}. \tag{3.22}$$

The Kalman gain K_k can be computed using standard formulas reproduced in the Appendix.

3.3.2. *The new law motion for the hierarchy of average expectations.* To find the law of motion for the average expectations from order 1 to $k+1$, take the cross-sectional average of the signal vector $\mathbf{s}_{t,j}$ and substitute into (3.19) to get

$$\theta_t^{(1:k+1)} = (I - K_k D_k) M_k \theta_{t-1}^{(1:k+1)} + K_k D_k M_k \theta_{t-1}^{(0:k+1)} + (K_k D_k N_k + K_k R_w) \mathbf{w}_t. \tag{3.23}$$

By amending this expression to the law of motion (2.2) of the actual process θ_t we get the law of motion for the hierarchy of average expectations from order zero to $k+1$ as

$$\begin{bmatrix} \theta_t \\ \theta_t^{(1:k+1)} \end{bmatrix} = M_{k+1} \begin{bmatrix} \theta_{t-1} \\ \theta_{t-1}^{(1:k+1)} \end{bmatrix} + N_{k+1} \mathbf{w}_t \tag{3.24}$$

where the matrices M_{k+1} and N_{k+1} are given by

$$M_{k+1} = \begin{bmatrix} \rho & \mathbf{0}_{1\times k} \\ \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times k} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{1\times k} & 0 \\ K_k D_k M_k & \mathbf{0}_{k\times 1} \end{bmatrix} + \begin{bmatrix} 0 & \mathbf{0}_{1\times k} \\ \mathbf{0}_{k\times 1} & (I - K_k D_k) M_k \end{bmatrix} \tag{3.25}$$

$$N_{k+1} = \begin{bmatrix} \sigma_v e_1' \\ (K_k D_k N_k + K_k R_w) \end{bmatrix}. \tag{3.26}$$

Taking a state space system of the form (3.18)-(3.21) with k orders of expectations as given, we can thus use the steps (3.22)-(3.26) to compute the law of motion of the state with $k+1$ orders of expectations.

3.3.3. *The higher-order expectations operator.* To complete a full step of the recursion, we also need to find the new price function p_t^{k+1} that depends on $k+1$ orders of expectations. To do so, it is useful to define the *higher-order expectations operator* $H_k : \mathbb{R}^{k+1} \to \mathbb{R}^k$ as

$$\theta_t^{(1:k)} = H_k \theta_t^{(0:k)}. \tag{3.27}$$

That is, the operator H applied to a hierarchy of expectations moves the hierarchy one step up in orders of expectations by annihilating the first element in the hierarchy.[4] It is given by the matrix

$$H_k \equiv \begin{bmatrix} \mathbf{0}_{k\times 1} & I_k \end{bmatrix}. \tag{3.28}$$

[4] Allen, Morris and Shin (2006) defines an *average belief operator* $\overline{E} : \mathbb{R}^2 \to \mathbb{R}^2$. The operator $\overline{E}$ maps the average k order expectations of the average signal vector into $k+1$ order expectations of the same vector and can be used to compute higher order expectations of the state since the static setting results in a proportional relationship between higher order beliefs. In our model, the elements of N_{k+1} in the law of motion (3.24) could be generated by a similar operator if θ_t was a non-persistent process.

3.3.4. *The new price function.* Multiplying a hierarchy of expectations by $M_k H_{k+1}$ moves the hierarchy one step up in orders of expectations and one step forward in time. The average expectation in period t of the price in period $t+1$ can then be computed as

$$\int E\left[p_{t+1}^k \mid \Omega_{t,j}^k\right] \, dj = \mathbf{g}_k' M_k H_{k+1} \theta_t^{(0:k+1)}. \tag{3.29}$$

Substituting this average expectation into the Euler equation (2.1) then gives the $k+1$ order price process as

$$p_t^{k+1} = \beta \mathbf{g}_k' M_k H_{k+1} \theta_t^{(0:k+1)} - \theta_t - \varepsilon_t. \tag{3.30}$$

The new price function can thus be expressed as

$$p_t^{k+1} = \mathbf{g}_{k+1}' \theta_t^{(0:k+1)} - \varepsilon_t \tag{3.31}$$

where

$$\mathbf{g}_{k+1}' = e_1' + \beta \mathbf{g}_k' M_k H_{k+1}. \tag{3.32}$$

The steps above define an algorithm that takes the price process (3.18)-(3.20) with k orders of expectations as an input and produces a new price process given by (3.24)-(3.30) with $k+1$ orders of expectations as an output.

4. Existence, Uniqueness and Approximation Error Bounds

The previous section derived a recursive algorithm that can be used to compute the law of motion for the hierarchy of expectations $\theta_t^{(0:k)}$ and the associated price function p_t^k for an arbitrarily large k. By itself, this algorithm thus does not solve the infinite regress problem. However, in this section we first show that there exists a unique solution to the model by proving that the mapping implied by the agents' filtering and prediction problem is a contraction. The recursive algorithm above is then shown to be a special case of this mapping. As a consequence, the sequence of price processes $p_t^k : k = 0, 1, 2, ...$ converges to the unique equilibrium of the model as k increases. From the contractive property of the algorithm, it also follows that any desired degree of solution accuracy can be achieved with a finite-dimensional state representation and we present explicit bounds on the approximation errors.

The structure of the formal results below is as follows. We first define the relevant space of analysis and then prove two intermediate lemmas. The first lemma formally defines the mapping implied by the Euler equation (2.1) and we prove that it is a self-map on the relevant Hilbert space. This mapping takes a given process for the price and maps that into a new price process by solving the agents' filtering and prediction problem, taking the initial price process as given. To show that it is a contraction mapping requires us to compare the distance between two arbitrary price processes with the distance of the price processes after applying the mapping. The second lemma facilitates this step by showing how the average expectations in the Euler equation (2.1) can be rewritten to make these distances easy to compute and compare.

It is helpful to have a notation that makes a distinction between particular realizations of a time series process and the time series process itself. We will use $x \equiv \{x_t\}_{t=-\infty}^{\infty}$ to denote the time series process, and x_t to denote the realized value of x in period t. The relevant space for the analysis is $\mathcal{L}^2$, i.e. the space of all random processes with a finite variance, which we now define.

Definition 1. *(The space $\mathcal{L}^2$) The real Hilbert space $\mathcal{L}^2$ is the collection of all random variables x with finite variance*

$$Ex^2 < \infty \tag{4.1}$$

endowed with the inner-product

$$\langle x, y\rangle \equiv E(xy) : x, y \in \mathcal{L}^2 \tag{4.2}$$

and induced norm

$$\|x\| = \sqrt{\langle x, x\rangle}. \tag{4.3}$$

The space $\mathcal{L}^2$ is a well-understood Hilbert space commonly used to study time series processes, see for instance Brockwell and Davis (2006). The norm (4.3) implies that the distance between two elements x and y in $\mathcal{L}^2$ is the standard deviation of the difference $x_t - y_t$. Two processes are equivalent if they are equal almost surely.

Proposition 1 below establishes that the recursive algorithm to compute the sequence of price processes $p^0, p^1, ..., p^k$, while increasing the number of orders of expectations that is taken into account, is a contraction mapping. As an intermediate step towards this end, we first show that mapping implied by the algorithm is a self-map on $\mathcal{L}^2$.

Lemma 1. *(The mapping $T : \mathcal{L}^2 \to \mathcal{L}^2$) For the time series process $x \in \mathcal{L}^2$, define the mapping T of x as*

$$Tx \equiv \beta \int E\left[x_{t+1} \mid \Omega^x_{t,j}\right] dj + \theta_t + \varepsilon_t. \tag{4.4}$$

where $\Omega^x_{t,j} \equiv \left\{x_t, z_{t,j}, \Omega^x_{t-1,j}\right\}$. Then T is a self-map on $\mathcal{L}^2$ so that $x \in \mathcal{L}^2$ implies that $Tx \in \mathcal{L}^2$.

Proof. In the Appendix. □

Note that the mapping from p^k to p^{k+1}, defined by the steps from (3.18) to (3.30) is of the same form as T. It takes the price process defined by (3.18) and (3.20) as given, then solves the implied filtering and prediction problem of an agent to compute the average expectation (3.29). Substituting this average expectation into (2.1) then corresponds to the operation (4.4) performed by T, resulting in the new price process given by (3.24) and (3.30).

The space $\mathcal{L}^2$ is closed under addition, so the proof of Lemma 1 simply entails showing that the variance of the average period t expectation of x_{t+1} is finite if $x \in \mathcal{L}^2$. The initial price process p^0, as defined by (3.6), has a finite variance and thus belongs to $\mathcal{L}^2$. By the principle of induction, Lemma 1 then holds for each step k of the algorithm so that $p^{k+1} = Tp^k \in \mathcal{L}^2$ for every k.

The main result of this section is to demonstrate that the sequence of price processes $p^k : k = 0, 1, 2, ...$ converges to the unique equilibrium of the model (2.1)-(2.4) by showing that T is a contraction on $\mathcal{L}^2$. To do so, it is helpful to first rewrite the average expectation in (4.4) in an equivalent form that simplifies computing the distance $\|Tx - Ty\|$ for $x, y \in \mathcal{L}^2$.

Lemma 2. *(MA representation of the average expectation) The average expectation* $\int E\left[x_{t+1} \mid \Omega^x_{t,j}\right] dj$ *can be written as*

$$\int E\left[x_{t+1} \mid \Omega^x_{t,j}\right] dj = e_2' \left[L^{-1}C_x(L)\right]_+ \mathbf{w}_t \tag{4.5}$$

where $C_x(L)$ *is a lag polynomial and* $[\quad]_+$ *is the annihilation operator that sets all negative powers of* L *to zero.*

Proof. From the Wold Decomposition Theorem (e.g. Brockwell and Davis 2006), we can write the process for agent j's signal vector $\mathbf{s}_{t,j}$ as

$$\mathbf{s}_{t,j} = A_x(L)\widetilde{\mathbf{s}}_{t,j} \tag{4.6}$$

where $A_x(L)$ is a causal lag polynomial in the white noise process $\widetilde{\mathbf{s}}_{t,j}$. The white noise process $\widetilde{\mathbf{s}}_{t,j}$ is defined as the innovation to agent j's signal vector

$$\widetilde{\mathbf{s}}_{t,j} \equiv \mathbf{s}_{t,j} - E\left[\mathbf{s}_{t,j} \mid \mathbf{s}_{t-1,j}, \mathbf{s}_{t-2,j}, ...\right]. \tag{4.7}$$

Because $\widetilde{\mathbf{s}}_{t,j} \in \Omega^x_{t,j}$, we can use the Wiener-Kolmogorov prediction formula (e.g. Hansen and Sargent 1981) to write agent j's expectation about the next period signal vector as

$$E\left[\mathbf{s}_{t+1,j} \mid \Omega^x_{t,j}\right] = \left[L^{-1}A_x(L)\right]_+ \widetilde{\mathbf{s}}_{t,j}. \tag{4.8}$$

To find agent j's expectation of x_{t+1}, use that x_t is a component of $\mathbf{s}_{t,j}$ so that

$$E\left[x_{t+1} \mid \Omega^x_{t,j}\right] = e_2' \left[L^{-1}A_x(L)\right]_+ \widetilde{\mathbf{s}}_{t,j}. \tag{4.9}$$

The last step to get the desired form (4.5) is to note that by (3.9) the innovation to the average signal vector can be written in the form

$$\int \widetilde{\mathbf{s}}_{t,j} dj = B_x(L)\mathbf{w}_t. \tag{4.10}$$

Integrating (4.9) over j and substituting in (4.10) for $\int \widetilde{\mathbf{s}}_{t,j} dj$ in (4.9) and defining

$$C_x(L) \equiv A_x(L)B_x(L) \tag{4.11}$$

gives the desired expression.[5,6] □

We now have the components needed to prove that the mapping T is a contraction, which is the main proposition of this section.

[5]While not needed for the proof, we can compute the coefficients in both $A(L)$ and $B(L)$ explicitly from the definition (3.21) of the signal vector and the Kalman filter update equation (3.19) so that $\mathbf{z}_{t,j} = \widetilde{\mathbf{z}}_{t,j} + L\Sigma_{s=1}^{\infty} M^s K \widetilde{\mathbf{z}}_{t-s,j}$ and $\int \widetilde{\mathbf{z}}_{t,j} dj = (LN + R)\,\mathbf{w}_t + L\Sigma_{s=1}^{\infty} M^s (N - KLN - KR)\,\mathbf{w}_{t-s}$.

[6]The lag polynomial $A(L)$ is invertible by construction, but $C(L)$ is not invertible unless agents' signals are perfectly revealing. Rondina and Walker (2014) propose a method that solves directly for the non-invertible MA representation of the endogenous variable. They specify their exogenous processes so that it is possible to "flip" what corresponds to $A(L)$ in their model into an observationally equivalent, but non-invertible MA representation. This non-invertible representation then corresponds to $C(L)$ here. However, their technique requires restrictive assumptions on the functional form of $A(L)$. For instance, it cannot be used to solve a model such as (2.1)-(2.4) in which agents observe private signals about an AR(1) process.

Proposition 1. *(T is a contraction) The mapping T is a contraction with contractive constant $0 \leq \beta < 1$.*

Proof. For any $x, y \in \mathcal{L}^2$, we need to show that

$$\|Tx - Ty\| \leq \beta \|x - y\| . \tag{4.12}$$

We start by finding the distance $\|x - y\|$.

Using Wold's theorem as in Lemma 2, we can write x and y in MA form as

$$x_t = e_2' C_x(L)\mathbf{w}_t, \quad y_t = e_2' C_y(L)\mathbf{w}_t.$$

or more explicitly as

$$x_t = e_2' \sum C_{x,s}\mathbf{w}_{t-s}, \quad y_t = e_2' \sum C_{y,s}\mathbf{w}_{t-s}.$$

The distance $\|x - y\|$ induced by the norm (4.3) is the standard deviation of $x_t - y_t$. Since $E\mathbf{w}_t\mathbf{w}_t' = I$ and $E\mathbf{w}_t\mathbf{w}_{t-s}' = \mathbf{0}$ for $s \neq 0$, it can be computed as

$$\|x - y\| = \sqrt{\sum_{s=0}^{\infty} \left(c_{2,1,x,s} - c_{2,1,y,s}\right)^2 + \sum_{s=0}^{\infty} \left(c_{2,2,x,s} - c_{2,2,y,s}\right)^2} \tag{4.13}$$

where $c_{i,j,x,s}$ and $c_{i,j,y,s}$ are the i, j element of $C_{x,s}$ and $C_{y,s}$ respectively.

The compare the distance $\|Tx - Ty\|$ with $\beta \|x - y\|$, use that we can express as $\|Tx - Ty\|$ as

$$\begin{aligned} \|Tx - Ty\| &= |\beta| \left\| e_2' \left[L^{-1}C_x(L)\right]_+ \mathbf{w}_t - e_2' \left[L^{-1}C_y(L)\right]_+ \mathbf{w}_t \right\| && (4.14) \\ &= |\beta| \sqrt{\sum_{s=1}^{\infty} \left(c_{2,1,x,s} - c_{2,1,y,s}\right)^2 + \sum_{s=1}^{\infty} \left(c_{2,2,x,s} - c_{2,2,y,s}\right)^2} && (4.15) \\ &\leq \beta \|x - y\| . && (4.16) \end{aligned}$$

The first equality follows from combining the definition (4.4) of T with Lemma 2. The second equality follows from the definition of the norm. The last inequality, which concludes the proof, follows from the fact that Wiener-Kolmogorov prediction formula $\left[L^{-1}C(L)\right]_+$ annihilates the first term in the expansion of $C(L)$. The sums in (4.15) thus contain one non-negative term less than the sums in (4.13). □

The fact that T is a contraction has several useful implications that we summarize in the following corollaries.

Corollary 1. *(T has a unique fixed point) There exists a unique $p^* \in \mathcal{L}^2$ such that*

$$p^* = Tp^*. \tag{4.17}$$

That T has a unique fixed point follows from standard results about contractive maps, see for instance Kreyszig (1978) or Atkinson and Han (2005). Note that the proof of Proposition 1 allows x and y to be any stationary linear time series process with Gaussian shocks. The proof thus do not rely on the particular recursive representation of the state proposed here. The fixed point of T is the unique equilibrium of the model when agents condition on the entire history of $\mathbf{s}_{t,j}$, regardless of how this equilibrium is represented.

Proposition 1 implies that the price process p^k will converge to the unique fixed point p^* as k increases. Because T is a contraction mapping with contractive constant β, we can derive explicit bounds on the approximation errors $p^k - p^*$ as a function of the number of iterations k.

Corollary 2. *(Convergence and error bounds) For any $p^0 \in \mathcal{L}^2$, the sequence $\left\{p^k\right\}_{k=0}^{\infty}$ defined by $p^{k+1} = Tp^k$ converges to $p^* \in \mathcal{L}^2$ and the following bounds are valid*

$$\left\|p^k - p^*\right\| \leq \frac{\beta^k}{1-\beta}\left\|p^0 - p^1\right\| \tag{4.18}$$

$$\left\|p^k - p^*\right\| \leq \frac{\beta}{1-\beta}\left\|p^{k-1} - p^k\right\| \tag{4.19}$$

Again, the proof follows from standard results about contractive maps. The contractive property of T implies that $\left\{p^k\right\}_{k=0}^{\infty}$ is a Cauchy sequence. From the completeness of $\mathcal{L}^2$ it then follows that $p^* \in \mathcal{L}^2$, i.e. that the equilibrium price has finite variance. Because $\left\|p^0 - p^1\right\|$ is finite and $|\beta| < 1$, the bound (4.18) implies that for any desired degree of precision $\delta > 0$, there exists a finite k such that

$$\left\|p^k - p^*\right\| < \delta. \tag{4.20}$$

After computing the first iteration on T, the bound (4.18) can be used to compute how many iterations are needed in total for the approximation error to be smaller than δ. The inequality (4.19) can be used to bound the error after having computed k iterations.

4.1. **A finite-dimensional approximation.** In the k^{th} iteration, the dimension of the state is $k+1$. From the bound (4.18) it follows that for any desired accuracy, we can approximate the equilibrium dynamics with a finite-dimensional state. This is useful, since previous results have suggested that there may be no exact finite-dimensional state representation of the equilibrium dynamics in models where privately informed agents extract information from endogenous variables. For instance, Makarov and Rytchkov (2012) shows that the equilibrium dynamics of a dynamic asset pricing model with two groups of privately informed traders cannot be represented by a finite order ARMA process. Huo and Takayama (2015a) prove *a no-finite-order-ARMA-representation-with-endogenous-signals* result in the same spirit as that of Makarov and Rytchkov's in the context of static decision model with a continuum of privately informed agents. That we can show here that a unique equilibrium exists and that it can be approximated arbitrarily well with a finite-dimensional state is thus an important result from a practical perspective since in many settings it is natural to assume that agents can observe some endogenous variables.[7] In the next section we demonstrate how to apply the solution method proposed here in practice and in Section 6 we generalize the results to a class of models that allow for multiple endogenous variables and stationary, but otherwise unrestricted, VARMA processes for the exogenous variables.

[7]Huo and Takayama (2015a) conjecture that the equilibrium dynamics of their model with endogenous signals can be approximated with a finite order ARMA process. However, they do not provide a proof of this conjecture.

5. Solving the asset pricing model

In this section we illustrate how the solution method works in practice by applying it to the model used in the analysis above. To do so, we first need to choose values for the parameters of the model. In the benchmark parameterization, we set $\{\beta, \rho, \sigma_v, \sigma_\varepsilon, \sigma_\eta\} = \{0.95, 0.9, 0.05, 1, 0.1\}$. We also need to decide how many iterations on the algorithm that are required to achieve a satisfactory accuracy of the solution. How small the approximation errors should be for the solution to be considered sufficiently accurate will generally depend on the application in question. For illustrative purposes, we here solve and simulate the model by computing 50 iterations on the algorithm.

The norm $\left\|p^k - p^*\right\|$ in the bounds (4.18) and (4.19) corresponds to the standard deviation of the approximation error. After 50 iterations on the algorithm, this bound, expressed as a fraction of the standard deviation of the price p^k, is 0.0006 or six hundredths of a percent of the standard deviation of the price.[8] The speed of convergence, and as a consequence, how many iterations that are needed to solve the model accurately, generally depends on the parameter values. Section 5.3 below studies this dependence in more detail. The number of iterations also equals the maximum order of expectations considered by the agents. Below, we denote this maximum order of expectation $\overline{k}$.

5.1. **Equilibrium price dynamics.** In the top row of Figure 1, we have plotted the impulse response function of the price of the asset to an innovation to the persistent component of supply (left column) and to a transitory shock (right column) using the benchmark parameterization. For comparison, we have also plotted the impulse response to the same innovation under the alternative assumptions of full information, i.e. when θ_t is observed perfectly by all agents.

It is clear from inspecting Figure 1 that the different information structures imply very different price dynamics. Noisy private signals result in weaker initial responses to a persistent supply shock compared to the full information case. Imperfect information also makes the price response to a transitory shock persistent somewhat persistent. That private information can be a strong force of inertia in endogenous variables has been noted before, e.g. in the macroeconomic applications in Woodford (2002), Nimark (2008), Graham and Wright (2010) and Angeletos and La'O (2009).

As first pointed out by Woodford (2002), the inertial response of the endogenous variable is caused by the sluggish response of higher-order expectations to a persistent shock. This is illustrated in the bottom row of Figure 1, where the responses of the hierarchy of expectations about θ_t to the two shocks are plotted. After an innovation to θ_t, first order expectations respond less on impact than the true shock. First order expectations also respond with inertia, while the true shock converges geometrically towards zero after the shock as implied by its AR(1) structure. This pattern is more pronounced for higher-order expectations, which respond less on impact and with more inertia than lower order expectations.

It is common knowledge of rational expectations that cause higher-order expectation to respond less than lower order expectations to persistent supply shocks. To see how, note that

[8] At $k = 100$, the corresponding approximation error-to-price standard deviation ratio is 2×10^{-7}. Computing 50 iterations takes about 0.25 seconds on a standard desktop computer. 100 iterations takes about 1 second.

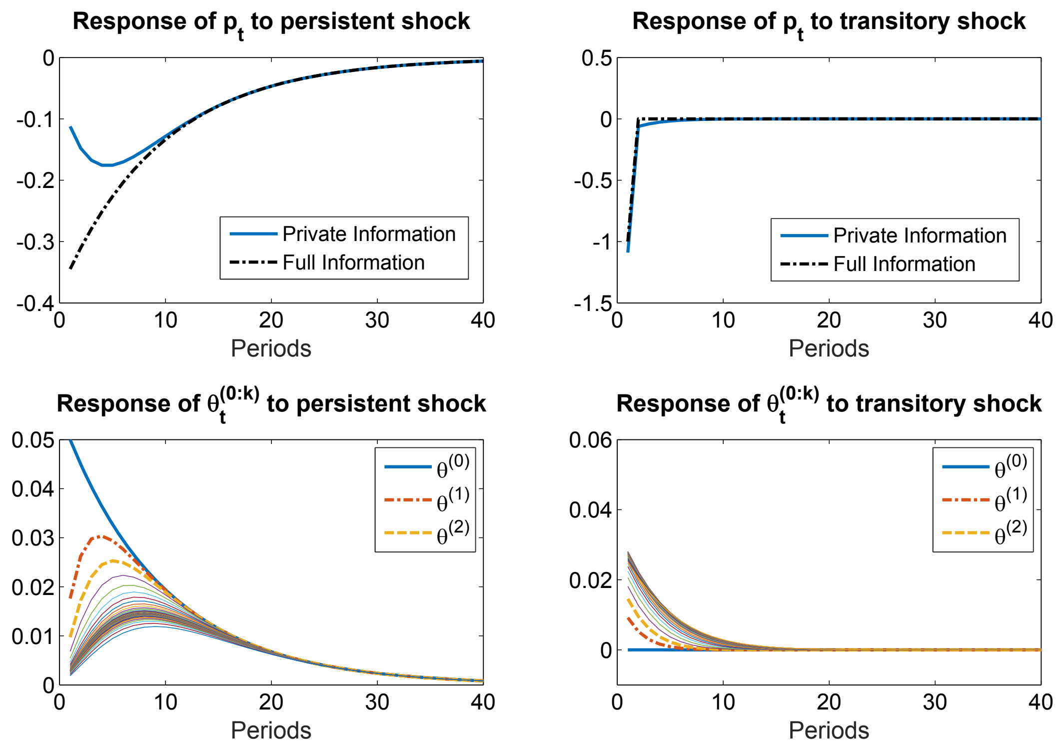

FIGURE 1. Impulse responses of p_t (top row) and $\theta_t^{(0:50)}$ (bottom row) to innovation to persistent (left column) and transitory (right column) component of supply.

first order expectation respond less than the true shock on impact because some of the price movement that is due to the persistent supply shock θ_t will be attributed to the transitory shock ε_t. Since agents know that first order expectations on average respond less than the actual shock, second order expectations must respond even less than first order expectations to the same shock. This argument can then be applied recursively to understand why a $k+1$ order expectation responds less than a k order expectation in the impact period.

After a transitory shock ε_t, expectations about θ_t respond with some persistence as it takes time for agents to realize that there has been no actual change in θ_t. In contrast to the response to an innovation to the persistent supply shock θ_t, higher-order expectations respond more strongly than lower order expectations to the transitory shock ε_t. However, as can be seen in the top right panel of Figure 1, the implied impact of the higher-order expectations about θ_t on p_t is small relative to the direct effect of ε_t and the response of p_t is close to what it would be under full information.

5.2. **Cross-sectional dispersion of expectations.** Because agents have access to private information, the cross-sectional dispersion of expectations in the model is non-degenerate. Unlike full information models, or models that assume that there are two distinct groups of agents such as in the paper by Makarov and Rytchkov (2012) or Kasa, Walker and Whiteman (2014), the model here can match the survey evidence that suggests that market participants have dispersed expectations about future economic outcomes, e.g. Thomas (1999), Mankiw, Reis and Wolfers (2003), Swanson (2006) and Coibion and Gorodnichenko (2012). The same survey evidence can also be used to discipline the parameters of a model with privately informed agents. This can be done formally through likelihood based estimation as in Nimark (2014) and Barillas and Nimark (2016b) who treat individual survey responses from the *Survey of Professional Forecasters* as being representative of the forecasts of agents drawn randomly from the model's population. A less formal approach is to compare the implied forecast dispersion of a calibrated model to the data on forecast disagreement. To follow either of these approaches, it is necessary to compute the model implied cross-sectional dispersion of expectations in the model and we now show how to do so in the context of the model.

In the model, the dispersion of expectations is driven by the idiosyncratic noise $\eta_{t,j}$ in the private signals $z_{t,j}$. The idiosyncratic noise shocks $\eta_{t,j}$ are white noise processes that are orthogonal across agents and to the supply shocks u_t and ε_t. This implies that the cross-sectional variance of expectations is equal to the part of the unconditional variance of agent j's expectations that is due to the idiosyncratic shocks. This quantity can be computed by finding the variance of the estimates in agent j's updating equation (3.19), but with the aggregate shocks u_t and ε_t "switched off".

Define the agent specific covariance Σ_j of agent j's state estimate as

$$\Sigma_j \equiv E\left(\theta_{t,j}^{(1:\overline{k})} - \int \theta_{t,j'}^{(1:\overline{k})} dj'\right)\left(\theta_{t,j}^{(1:\overline{k})} - \int \theta_{t,j'}^{(1:\overline{k})} dj'\right)'. \tag{5.1}$$

The agent specific covariance can then be computed by the solving the Lyaponov equation

$$\Sigma_j = \left(I - K_{\overline{k}} L_{\overline{k}}\right) M_{\overline{k}} \Sigma_j M_{\overline{k}}' \left(I - K_{\overline{k}} D_{\overline{k}}\right)' + K_{\overline{k}} R_\eta R_\eta' K_{\overline{k}}'. \tag{5.2}$$

The covariance Σ_j captures cross-sectional dispersion of expectations about the state $\theta_t^{(0:\overline{k}-1)}$. What we observe in the survey data is generally forecasts about endogenous variables, such as future interest rates or future inflation. Given the solved model, it is straight-forward to translate the dispersion of expectations about the latent state into dispersion about forecasts of endogenous variables. For instance, to compute the model implied cross-sectional variance $\sigma_{p,j}^2$ of expectations about the next period price p_{t+1}, we can use the price function (3.30) to get

$$\sigma_{p,j}^2 = \mathbf{g}_{\overline{k}}' M_{\overline{k}} \Sigma_j M_{\overline{k}}' \mathbf{g}_{\overline{k}}. \tag{5.3}$$

In the parameterization used to solve the model here, the cross-sectional standard deviation of one period ahead price forecasts is 0.15, or about 12% of the standard deviation of the price.

5.3. **Accuracy and convergence.** The dynamics of the endogenous price p_t is completely summarized by its MA representation. Figure 2 illustrates the convergence properties of the algorithm in this space by plotting the MA representation of the price for each iteration $k = 1, 2, ..., \overline{k}$. After 20 iterations, the algorithm has for all practical purposes converged and the MA representations for iterations $k > 20$ are indistinguishable from each other. This may suggest that about 20 orders of expectation matter quantitatively in equilibrium. However, this is not the case. Most of the changes in the MA representations that occur as

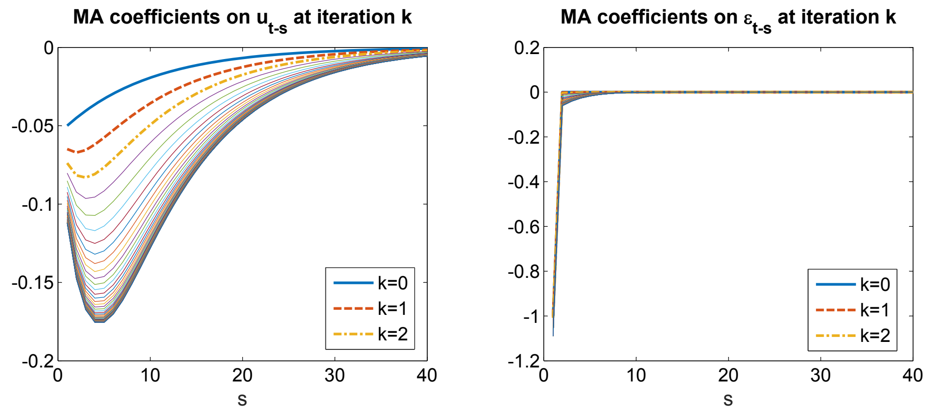

FIGURE 2. MA representation of p_t^k for $k = 1, 2, ..., \overline{k}$. Thin solid lines denote the MA representation for $k > 2$.

k increases are not due to the inclusion of increasingly high orders of expectations. Instead, most of the changes are caused by an accumulation of weights attached to relatively low orders of expectations in the price function (3.30). This is illustrated in Figure 3, where we have plotted the loadings in the price function $\mathbf{g}_k'$ for $k = 0, 1, 2, ...$ After 50 iterations, most of the weight in the price function loads on expectations of order 6 or less, with the largest weight attached to the average first order expectation. Most of the changes in the MA representation in Figure 2 as k increases is thus due to the increasing weight attached to lower order expectations.

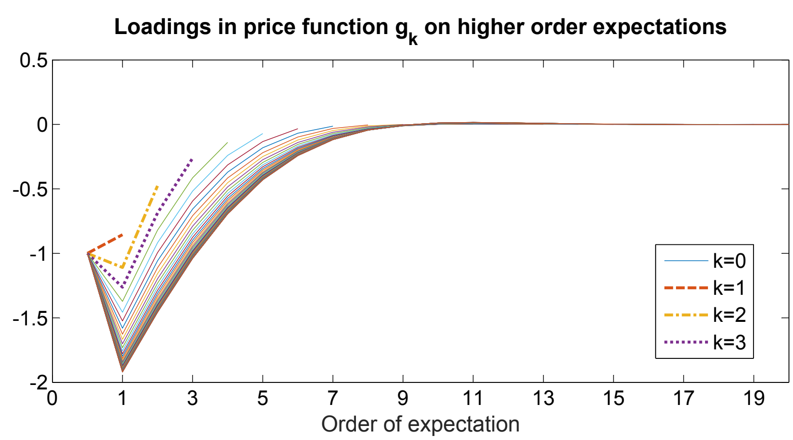

FIGURE 3. Price function loadings on the k order (x-axis) expectation of price, i.e. the elements of $\mathbf{g}_k$ for $k = 1, 2, ..., \overline{k}$. The increasing dimension of the state is manifested by the increasing dimension of the vector $\mathbf{g}_k$. Note that $\mathbf{g}_0$ is represented by a point at $(0, -1)$.

The fact that most of the weight in the price function loads on relatively low orders of expectations suggests that only a few orders of expectations matter quantitatively. A second force working in the same direction is the fact that the variance of higher-order expectations is bounded by the variance of lower order expectations. So not only do higher-order expectations have less impact on the price, they also have lower variance. Together, these two features imply that once we have found an accurate law of motion for the state with 50 orders of expectations, we can in practice set $\theta_t^{(k)} = 0$ for $k > 6$ without significantly altering the dynamics of p_t. The standard deviation of the difference between the price process with only the first six orders of expectations and the price process using 50 orders of expectations is 0.0007. For perspective, this can be compared to the standard deviation of $p^{\overline{k}}$ which is equal to 1.23. While not shown in the figures, to the naked eye, the implied MA representation of using only the first six orders of expectations is indistinguishable from the MA representation of $p^{\overline{k}}$. The equilibrium of the simple asset pricing model presented here can thus be accurately represented by a low-dimensional state vector.

5.4. **Parameters and approximation error bounds.** As mentioned at the beginning of this section, the number of iterations required to reach an accurate solution generally depends on the parameters of the model. Figure 4 illustrates how the size and convergence rates of the approximation error bounds varies with different parameterizations. In each panel of the figure, the units on the x-axis is the number of iterations and units on the y-axis is the ratio of the standard deviation of the approximation error and the standard deviation of the price at the equilibrium fixed point.

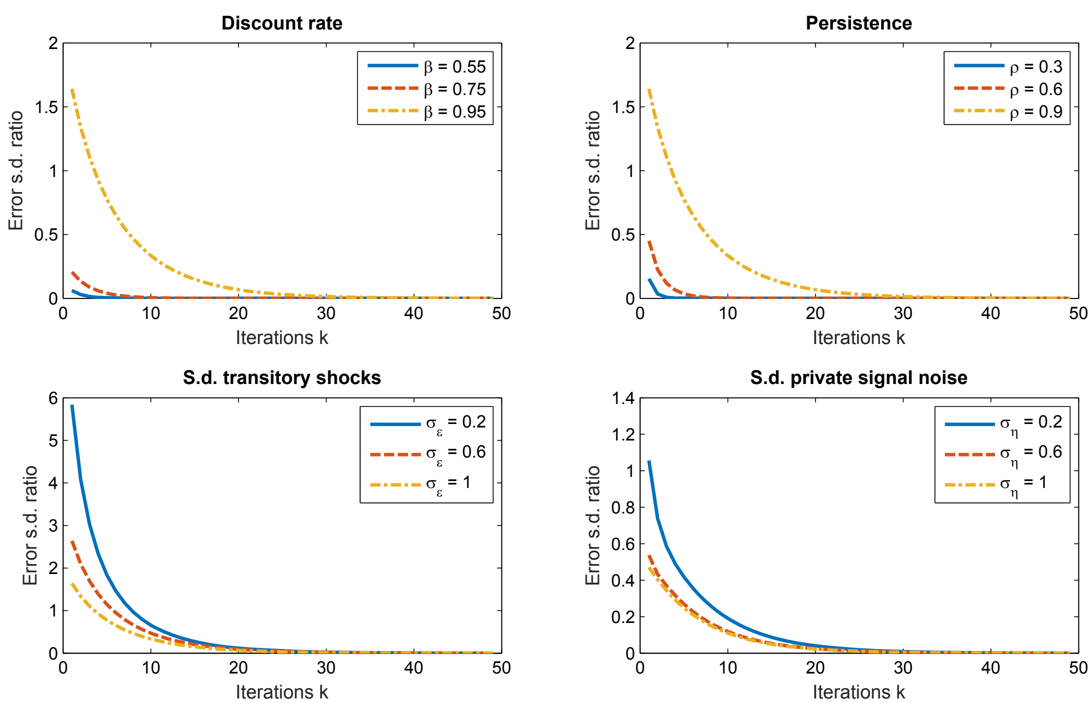

FIGURE 4. Error bound dependence on number of iterations and (selected) parameter values.

The top left panel shows how the error bound (4.19) depends on the discount rate β. Unsurprisingly, given that β is also the contractive constant of the algorithm, the lower β is, the smaller are the approximation errors and the faster is the convergence of the error bound towards zero. The top right panel shows that a similar relationship holds for the parameter that governs the persistence of θ_t. The more persistent this process is, the larger is the number of iterations needed to achieve a given level of accuracy.

The two bottom panels in Figure 4 shows how the accuracy of the solution depends on parameters that determine how precise agents' information is. Lower standard deviations of either the transitory supply shock ε_t or the idiosyncratic noise shocks $\eta_{t,j}$ imply that agents have more precise estimates of θ_t. Lower standard deviations of these shocks also imply that more iterations are needed to achieve a given degree of accuracy. A limit case may provide some intuition for why this is the case. As signals become infinitely noisy, agents will disregard them completely and expectations will not respond at all to shocks. Only the actual process for θ_t will then affect the price and the model is already at the fixed point at iteration $k = 0$.

Because of the contractive property of the algorithm, the exact specification of the initial price process p^0 does not matter for the equilibrium dynamics of the solved model. However, the required number of iterations for a given level of accuracy may depend on how the algorithm is initialized. In particular applications, there may exist numerically more efficient starting points than the process described by (3.6)-(3.9), in the sense that the algorithm may converge faster from other starting values. For instance, if one is primarily interested only in parameterizations with relatively precise signals, it may be better to start the algorithm from the full information solution (2.9).

6. A class of linear models with private information

This section generalizes the results above to a class of linear dynamic models with privately informed agents. Here we provide conditions that ensure the existence of a unique equilibrium in models with multiple endogenous variables and with stationary but otherwise unrestricted exogenous VARMA processes.

6.1. **A general model structure.** Consider the class of models that can be described by a vector valued Euler equation of the form

$$\mathbf{p}_t = \Lambda \int E\left[\mathbf{p}_{t+1} \mid \Omega_{t,j}\right] dj + F_\Theta \mathbf{\Theta}_t + F_w \mathbf{w}_t : \mathbf{w}_t \sim N(0, I_m) \tag{6.1}$$

where $\mathbf{p}_t$ is an n-dimensional vector of endogenous variables, $\mathbf{\Theta}_t$ is a q-dimensional vector of exogenous variables that follows a first order VAR process

$$\mathbf{\Theta}_t = M_0 \mathbf{\Theta}_{t-1} + N_0 \mathbf{w}_t \tag{6.2}$$

$\Lambda, F_{\mathbf{\Theta}}$ and F_w are matrices of conformable dimensions. $\Omega_{t,j}$ is the information set of agent j, defined by the filtration

$$\Omega_{t,j} = \left\{\mathbf{z}_{t,j}, \mathbf{p}_t, \Omega_{t-1,j}\right\}. \tag{6.3}$$

The vector of private signals $\mathbf{z}_{t,j}$ is of the form

$$\mathbf{z}_{t,j} = D_\Theta \mathbf{\Theta}_t + R_{zw} \mathbf{w}_t + R_{z\eta} \boldsymbol{\eta}_{t,j} : \boldsymbol{\eta}_{t,j} \sim N(0, I) \tag{6.4}$$

where $\boldsymbol{\eta}_{t,j}$ is a vector of agent j specific shocks.

The assumption that $\mathbf{\Theta}_t$ follows a first order VAR is non-restrictive in practice, as any finite order VARMA process always can be reformulated as an equivalent first order VAR. The model in Section 2 above is thus a special case of the more general form (6.1)-(6.4) which also nests the macro models of Nimark (2008), Melosi (2014) and Rondina and Walker (2014) and the bond pricing models of Nimark and Barillas (2016a, 2016b) and Struby (2016). To extend the results of Section 4 to this more general class of models, we first need to extend the space of analysis to allow for n endogenous variables.

Definition 2. (*The Banach space* $\mathcal{B}^n$) *The Banach space* $\mathcal{B}^n$ *is the* n^{th} *Cartesian power of* $\mathcal{L}^2$ *so that* $\mathcal{B}^n \equiv \underbrace{\mathcal{L}^2 \times \mathcal{L}^2 \times ... \times \mathcal{L}^2}_{n}$ *with associated norm* $\|\cdot\|_{\mathcal{B}}$ *defined as*

$$\|\mathbf{x}\|_{\mathcal{B}} \equiv \sum_{i=1}^{n} \|x_i\| : \mathbf{x} \in \mathcal{B}^n \tag{6.5}$$

where $\|\cdot\|$ *is the norm of* $\mathcal{L}^2$.

That the product space $\mathcal{B}^n$ is complete, and thus a Banach space, follows from the fact that $\|x_i\|$ is a norm on the (complete) Hilbert space $\mathcal{L}^2$ and because the norm $\|\cdot\|_{\mathcal{B}}$ on $\mathcal{B}^n$ is a conserving norm on the product space $\mathcal{B}^n$ (see Theorem 1.6.1 in O'Searcoid 2006).

With the natural modifications required to replace scalars with vectors and matrices, Theorem 1 extends the results of Section 4.

Theorem 1. *(Unique equilibrium and finite approximation) A model of the form (6.1)-(6.4) has a unique solution that can be approximated arbitrarily well with a finite-dimensional state if*

$$|eig(M_0)| < 1 \tag{6.6}$$

and

$$\|\Lambda\|_{\mathcal{B}} < \alpha \tag{6.7}$$

for some $0 \leq \alpha < 1$ *where* $\|\cdot\|_{\mathcal{B}}$ *is the operator norm on* $\mathcal{B}^n$ *defined as*

$$\|\Lambda\|_{\mathcal{B}} \equiv \sup_{x \neq 0} \frac{\|\Lambda \mathbf{x}\|_{\mathcal{B}}}{\|\mathbf{x}\|_{\mathcal{B}}}. \tag{6.8}$$

Proof. See Appendix. □

The logic of the proof of Theorem 1 follows closely that of Proposition 1, with the discounting condition $0 \leq \beta < 1$ replaced by the bound (6.7) on the matrix norm of Λ. The role of the eigenvalue bound (6.6) is to ensure that $F_\Theta \mathbf{\Theta}_t \in \mathcal{B}^n$ and corresponds to the condition that $0 \leq |\rho| < 1$ in the simple model of Section 2. The Appendix describes a modified version of the algorithm derived in Section 3 that can be used to solve a model of the form (6.1)-(6.4) taking the matrices $\Lambda, F_\Theta, F_w, M_0, N_0, D_\Theta, R_{zw}$ and $R_{z\eta}$ as inputs. The modified algorithm is again a contraction, implying that the equilibrium dynamics of $\mathbf{p}_t$ can be approximated to an arbitrary accuracy with a finite-dimensional state representation.

Both the proof of Proposition 1 and Theorem 1 explicitly use that agents observe the endogenous variables in $\mathbf{p}_t$. However, proving that the mapping T is a contraction is substantially simpler when agents observe only exogenous signals. The Appendix contains a proof that T is a contraction also when $\Omega_{t,j} = \{\mathbf{z}_{t,j}, \Omega_{t-1,j}\}$.

7. Conclusions

In many market settings, it is natural to assume that agents have access to private information that may influence their decisions and strategic behaviour. Yet, fully dynamic infinite horizon models with privately informed agents are arguably under-studied in the literature. A principal difficulty in studying this class of models is due to the infinite regress of expectations problem that arises from agents' need to "forecast the forecasts of others". In order to make progress, existing studies have been forced to make restrictive assumptions to circumvent the infinite regress of expectations.

This paper provides methods that can be used to analyze a relatively general class of linear dynamic rational expectations models that do feature an infinite regress of expectations. The paper makes two main contributions. First, we derived conditions that guarantee the existence of a unique equilibrium. The conditions for this result to hold are quite weak: It is sufficient that agents discount the future and that the exogenous processes follow stationary but otherwise unrestricted VARMA processes. Second, we showed that under the same conditions, and in spite of the infinite regress of expectations, the equilibrium of this class of models can be approximated arbitrarily well with a finite-dimensional state and explicit approximation error bounds.

The theoretical literature has to date produced a wealth of qualitative results derived from the interactions between privately informed agents. Many of these results have been derived using highly stylized models that provide sharp theoretical insights. Because the methods proposed here allow for a more general model structure with fewer restrictive assumptions than the existing alternatives, it can be used to solve richer models that more closely resembles those of the full information literature. The method proposed here thus helps shorten the step between the theoretical literature and the quantitative macroeconomics and finance literature. The solution method is also fast and flexible enough to estimate model parameters using likelihood-based methods, as evidenced by the applications in Nimark (2014), Melosi (2016), Barillas and Nimark (2016a, 2016b) and Struby (2016).

References

[1] Allen, F. S. Morris and H.S. Shin, 2006, "Beauty Contests and Iterated Expectations in Asset Markets", *Review of Financial Studies*, 19, pp719-752.

[2] Angeletos, G-M, and A. Pavan, 2007, "Efficient Use of Information and Social Value of Information", *Econometrica*, vol. 75(4), pp1103-1142.

[3] Angeletos, G-M, J. La'O, 2009, "Noisy Business Cycles", NBER Macroeconomics Annual, Cambridge, MA.

[4] Atkinson, K. and Han, W., 2005, *Theoretical numerical analysis*, Springer-Verlag.

[5] Bacchetta, Phillippe and Eric van Wincoop, 2006 "Can Information Heterogeneity Explain the Exchange Rate Determination Puzzle?", *American Economic Review* vol 96, pp552-576.

[6] Banerjee, S., Kaniel, R. and Kremer, I., 2009, "Price drift as an outcome of differences in higher-order beliefs", *Review of Financial Studies*, pp3707-3734.

[7] Barillas, F and K.P. Nimark, 2016a, "Speculation and the Term Structure of Interest Rates", forthcoming, *Review of Financial Studies.*

[8] Barillas, F and K.P. Nimark, 2016b, "Speculation and the Bond Market: An Empirical No-arbitrage Framework", *CEPR Discussion Paper* No. DP10892.

[9] Brockwell, P.J. and R.A. Davis, 2006, *Time Series: Theory and Methods*, Springer-Verlag.

[10] Cespa, Giovanni and Xavier Vives, 2012,"Dynamic trading and asset prices: Keynes vs. Hayek", *Review of Economic Studies*, 79(2), pp.539-580.

[11] Coibion, O. and Y. Gorodnichenko, 2012, "What can survey forecasts tell us about informational rigidities?", *Journal of Political Economy* pp116-159.

[12] Graham, L. and S. Wright, 2010, "Information, heterogeneity and market incompleteness", *Journal of Monetary Economics*, vol. 57(2), pp64-174.

[13] Grisse, C., 2009, "Higher Order Beliefs and the Comovement of Asset Prices", Working Paper, Federal Reserve Bank of New York.

[14] Hansen, L.P. and T.J. Sargent, 1981, "A note on Wiener-Kolmogorov prediction formulas for rational expectations models", *Economics Letters*, vol 8, pp255-260.

[15] Harsanyi, John C., 1967-68, "Games with Incomplete Information Played by Bayesian Players.' *Management Science* 14, pp159-182, pp320-334, and pp486-502.

[16] Hellwig, Christian , 2002. "Public Announcements, Adjustment Delays, and the Business Cycle (November 2002)," UCLA Economics Online Papers 208, UCLA Department of Economics.

[17] Hellwig, C. and V. Venkateswaran, 2009, "Setting the Right Price for the Wrong Reasons", *Journal of Monetary Economics*, vol. 56(S), ppS57-S77.

[18] Huo, Z. and Naoki Takayama, 2015a, "Rational Expectations Models with Higher Order Beliefs", mimeo, Yale University.

[19] Huo, Z. and Naoki Takayama, 2015b, "Higher Order Beliefs, Confidence, and Business Cycles", mimeo, Yale University.

[20] Kasa, K., T.B. Walker and C.H. Whiteman, 2014, "Heterogeneous beliefs and tests of present value models", *Review of Economic Studies*, pp1137-1163.

[21] Keynes, J.M., 1936, "The General Theory of Employment, Interest and Money". London, Macmillan.

[22] Kreyszig, E., 1978, *Introductory functional analysis with applications*, Wiley.

[23] Ljungqvist, L. and T.J. Sargent, 2004, "Recursive Macroeconomic Theory", MIT Press.

[24] Lorenzoni, Guido, 2009, "A Theory of Demand Shocks", *American Economic Review*, vol. 99(5), pp2050-84.

[25] Lorenzoni, Guido, 2010, "Optimal monetary policy with uncertain fundamentals and dispersed information", *Review of Economic Studies*, pp305-338.

[26] Lucas, Robert E., 1972, "Expectations and the Neutrality of Money", *Journal of Economic Theory*, pp103-124.

[27] Lucas, Robert E., 1973, "Some international evidence on output-inflation tradeoffs", *American Economic Review,* pp326-334.

[28] Lucas Jr, R.E., 1975, "An equilibrium model of the business cycle", *Journal of Political Economy*, pp1113-1144.

[29] Makarov, I. and O. Rytchkov, 2012, "Forecasting the forecasts of others: Implications for asset pricing", *Journal of Economic Theory,* pp941-966.

[30] Mankiw, G. R. Reis and J. Wolfers, 2003, "Disagreement about Inflation Expectations", NBER Macroeconomics Annual, 18, pp209-248.

[31] MaÄ‡kowiak, B. and Wiederholt, M., 2009. Optimal sticky prices under rational inattention. The American Economic Review, 99(3), pp.769-803.

[32] Melosi, Leonardo, 2016, "Signaling Effects of Monetary Policy", forthcoming, *Review of Economic Studies.*

[33] Morris, S. and H.S. Shin, 2006, "Inertia of Forward-Looking Expectations," *American Economic Review*, vol. 96(2), pp152-157.

[34] Nagel, Rosemarie., 1995, "Unraveling in Guessing Games: An Experimental Study", *American Economic Review* 85, no. 5 1313-326.
[35] Nimark, K., 2008. Dynamic pricing and imperfect common knowledge. Journal of Monetary Economics, 55(2), pp.365-382.
[36] Nimark, Kristoffer P., 2014, "Man-bites-dog Business Cycles", *American Economic Review*, 104(8), pp.2320-2367.
[37] Nimark, Kristoffer P., 2015, "A Low Dimensional Kalman Filter for Systems with Lagged States in the Measurement Equation", *Economic Letters*, pp10-13.
[38] O'Searcoid, Mícheál, 2006, *Metric spaces,* Springer Science & Business Media.
[39] Paciello, L. and M. Wiederholt, 2013, "Exogenous Information, Endogenous Information, and Optimal Monetary Policy", *Review of Economic Studies,* pp1-33.
[40] Phelps, Edmund S., 1970, "Introduction: The New Microeconomics in Employment and Inflation Theory," in E.S. Phelps et al., *Microeconomic Foundations of Employment and Inflation Theory*, New York: Norton, 1970.
[41] Rondina, G. and T. Walker, 2014, "Information Equilibria in Dynamic Economies", *mimeo*, UCSD and University of Indiana.
[42] Sargent, Thomas J., 1991, "Equilibrium with Signal Extraction from Endogenous Variables", *Journal of Economic Dynamics and Control* 15, pp245-273.
[43] Singleton, Kenneth J., 1987, "Asset prices in a time series model with disparately informed, competitive traders", in *New Approaches to Monetary Economics,* Eds. W.A. Burnett and K.J. Singleton, Cambridge University Press.
[44] Struby, Ethan, 2016,"Macroeconomic Disagreement in Treasury Yields", mimeo, Boston College.
[45] Swanson, E.T., 2006, "Have increases in Federal Reserve transparency improved private sector interest rate forecasts?", *Journal of Money, Credit, and Banking*, 38(3), pp.791-819.
[46] Thomas, L.B., 1999. "Survey measures of expected US inflation", *Journal of Economic Perspectives*, pp.125-144.
[47] Townsend, Robert M., 1983, Forecasting the Forecasts of Others, *Journal of Political Economy*, vol 91, pp546-588.
[48] Walker, Todd, 2007, "How equilibrium prices reveal information in a time series model with disparately informed, competitive traders", *Journal of Economic Theory*, pp512-537.
[49] Weinstein, J. and Yildiz, M., 2007, "Impact of higher-order uncertainty", *Games and Economic Behavior*, pp200-212.
[50] Woodford, M. 2002, "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.

Appendix A. The Kalman gain

The Kalman gain K_k associated with the system (3.20)-(3.21) can be computed using the formulas

$$\begin{aligned} K_k &= (\Sigma_k D_k' + N_k R')(D_k \Sigma_k D_k' + RR')^{-1} && \text{(A.1)} \\ \Sigma_k &= M_k \left(\Sigma_k - (\Sigma_k D_k' + N_k R')(D_k \Sigma_k D_k' + RR')^{-1} (\Sigma_k D_k' + N_k R')'\right) M_k' && \text{(A.2)} \\ &\quad + N_k N_k' \end{aligned}$$

where $R \equiv \begin{bmatrix} R_w & R_\eta \end{bmatrix}$. For a derivation, see for instance Ljungqvist and Sargent (2004) or Nimark (2015).

Appendix B. Proof of Lemma 1

Lemma 1 *(The mapping $T : \mathcal{L}^2 \to \mathcal{L}^2$) For the time series process $x \in \mathcal{L}^2$, define the mapping T of x as*

$$Tx \equiv \beta \int E\left[x_{t+1} \mid \Omega^x_{t,j}\right] dj + \theta_t + \varepsilon_t. \tag{B.1}$$

where $\Omega^x_{t,j} = \left\{x_t, z_{t,j}, \Omega^x_{t-1,j}\right\}$. Then T is a self map on $\mathcal{L}^2$ so that $x \in \mathcal{L}^2$ implies that $Tx \in \mathcal{L}^2$.

Proof. We want to show that when x belongs to $\mathcal{L}^2$, then so does Tx. The space $\mathcal{L}^2$ is closed under addition, and by definition, $\theta_t, \varepsilon_t \in \mathcal{L}^2$. It thus suffices to show that $x \in \mathcal{L}^2$ implies that $\beta \int E\left[x_{t+1} \mid \Omega^x_{t,j}\right] dj \in \mathcal{L}^2$, or equivalently, that $\left\|\beta \int E\left[x_{t+1} \mid \Omega^x_{t,j}\right] dj\right\| < \infty$.

The following steps deliver the desired result

$$\begin{aligned} \left\|\beta \int E\left[x_{t+1} \mid \Omega^x_{t,j}\right] dj\right\| &= |\beta| \left\|\Gamma_x(L)\mathbf{w}_t + \int \Lambda_x(L)\eta_{t,j} dj\right\| && \text{(B.2)} \\ &= |\beta| \left\|\Gamma_x(L)\mathbf{w}_t\right\| && \text{(B.3)} \\ &\leq |\beta| \left\|\Gamma_x(L)\mathbf{w}_t + \Lambda_x(L)\eta_{t,j}\right\| && \text{(B.4)} \\ &= |\beta| \left\|\mathcal{P}_{t,j} x_{t+1}\right\| && \text{(B.5)} \\ &\leq \left\|x_{t+1}\right\| && \text{(B.6)} \\ &< \infty && \text{(B.7)} \end{aligned}$$

The first equality follows from that the expectation $E\left[x_{t+1} \mid \Omega^x_{t,j}\right]$ lies in the space spanned by the current and lagged values of $\mathbf{w}_t$ and $\eta_{t,j}$ and thus can be expressed as $\Gamma_x(L)\mathbf{w}_t + \Lambda_x(L)\eta_{t,j}$. The equality on the second line follows from that $\int \Lambda_x(L)\eta_{t,j} dj = 0$. The inequality on the third line follows from the fact that $\mathbf{w}_t$ and $\eta_{t,j}$ are orthogonal. The equality on the fourth line follows from the projection theorem, which states that the conditional expectation $E\left[x_{t+1} \mid \Omega^x_{t,j}\right]$ exists and is equal to the projection $\mathcal{P}_{t,j} x_{t+1}$ of x_{t+1} onto the subspace spanned by $\Omega^x_{t,j}$ (e.g. Brockwell and Davis 2006). The inequality on the fifth line follows from that the operator norm of any projection is 1 so that $\|\mathcal{P}_{t,j} x_{t+1}\| \leq \|x_{t+1}\|$ and because $|\beta| < 1$. The last inequality follows from $x \in \mathcal{L}^2$. □

Appendix C. Proof of Theorem 1

The proof of Theorem 1 follows closely the structure of the proof of Proposition 1, with the natural adjustments necessary to allow for $\mathbf{p}_t \in \mathcal{B}^n$. We present the proof as a consequence of three separate lemmas, building on the results in Section 4. The first lemma shows that the mapping implied by the Euler equation (6.1) is a self-map on $\mathcal{B}^n$.

Lemma 3. *(The mapping $T : \mathcal{B}^n \to \mathcal{B}^n$) For the vector time series process $\mathbf{x} \in \mathcal{B}^n$, define the mapping T of $\mathbf{x}$ as*

$$T\mathbf{x} = \Lambda \int E\left[\mathbf{x}_{t+1} \mid \Omega^x_{t,j}\right] dj + F_\Theta \mathbf{\Theta}_t + F_w \mathbf{w}_t \tag{C.1}$$

where $\Omega^x_{t,j} \equiv \{\mathbf{x}_t, \mathbf{z}_{t,j}, \Omega_{t-1,j}\}$. Then T is a self-map on $\mathcal{B}^n$ so that $\mathbf{x} \in \mathcal{B}^n$ implies that $T\mathbf{x} \in \mathcal{B}^n$.

The eigenvalue bound (6.6) ensures that $F_\Theta \mathbf{\Theta}_t \in \mathcal{B}^n$. The proof of Lemma 3 then simply entails repeating the steps in the proof of Lemma 1, for each element $x_i : i = 1, 2, ...n$ where x_i is the i^{th} element of $\mathbf{x}$, establishing that if $\mathbf{x}$ is an element of $\mathcal{B}^n$, then so is $\int E\left[\mathbf{x}_{t+1} \mid \Omega^x_{t,j}\right] dj$.

The next lemma establishes that the average expectation $\int E\left[\mathbf{x}_{t+1} \mid \Omega^x_{t,j}\right] dj$ in (C.1) can be expressed in a form that makes it easier to compare the distances $\|T\mathbf{x} - T\mathbf{y}\|_{\mathcal{B}}$ and $\|\mathbf{x} - \mathbf{y}\|_{\mathcal{B}}$.

Lemma 4. *(MA representation of the average expectation) Define the matrix Q so that $\mathbf{x}_t = Q\mathbf{s}_{t,j}$. The average expectation $\int E\left[\mathbf{x}_{t+1} \mid \Omega^x_{t,j}\right] dj$ can be written as*

$$\int E\left[\mathbf{x}_{t+1} \mid \Omega^x_{t,j}\right] dj = Q\left[L^{-1} C_x(L)\right]_+ \mathbf{w}_t \tag{C.2}$$

where $C_x(L)$ is a lag polynomial and $[\quad]_+$ is the annihilation operator that sets all negative powers of L to zero.

Proof. Denote the vector of signals observed by agent j as $\mathbf{s}_{t,j}$. The proof then follows from steps (4.6)-(4.11) from Lemma 2, but with the vector e_2' replaced by the matrix Q. $\square$

The final lemma then uses the same logic as in Proposition 1, but generalized to vector valued processes in $\mathcal{B}^n$ to show that T is a contraction.

Lemma 5. *(T is a contraction) The mapping $T : \mathcal{B}^n \to \mathcal{B}^n$ is a contraction if*

$$\|\Lambda\|_{\mathcal{B}} < \alpha \tag{C.3}$$

for some $0 \leq \alpha < 1$.

Proof. For any $\mathbf{x}, \mathbf{y} \in \mathcal{B}^n$, we need to show that

$$\|T\mathbf{x} - T\mathbf{y}\|_{\mathcal{B}} \leq \alpha \|\mathbf{x} - \mathbf{y}\|_{\mathcal{B}}. \tag{C.4}$$

for some $0 \leq \alpha < 1$. We start by finding the distance $\|\mathbf{x} - \mathbf{y}\|_{\mathcal{B}}$.

Using Wold's theorem as in Lemma 4, we can write x and y in MA form as

$$\mathbf{x}_t = QC_x(L)\mathbf{w}_t, \quad \mathbf{y}_t = QC_y(L)\mathbf{w}_t. \tag{C.5}$$

or

$$\mathbf{x}_t = \sum \Gamma_{x,s}\mathbf{w}_{t-s}, \quad \mathbf{y}_t = \sum \Gamma_{y,s}\mathbf{w}_{t-s}. \tag{C.6}$$

where $\Gamma_{x,s} = QC_{x,s}$ and $\Gamma_{y,s} = QC_{y,s}$. The distance $\|\mathbf{x}-\mathbf{y}\|_{\mathcal{B}}$ induced by the norm (6.5) can then be computed as

$$\|\mathbf{x}-\mathbf{y}\|_{\mathcal{B}} = \sum_{i=1}^{n} \|x_i - y_i\|. \tag{C.7}$$

The components of the sum on the right hand side of (C.7) are given by

$$\|x_i - y_i\| = \sqrt{\sum_{\tau=1}^{m}\sum_{s=0}^{\infty}\left(\gamma_{i,\tau,x,s} - \gamma_{2,\tau,y,s}\right)^2} \tag{C.8}$$

where $\gamma_{i,\tau,x,s}$ is the i,τ element of $\Gamma_{x,s}$, $\gamma_{i,\tau,y,s}$ is the i,τ element of $\Gamma_{y,s}$ and m is the dimension of $\mathbf{w}_t$.

The compare the distance $\|T\mathbf{x}-T\mathbf{y}\|_{\mathcal{B}}$ with $\|\mathbf{x}-\mathbf{y}\|_{\mathcal{B}}$, use that from Lemma 4, and using the notation defined in (C.6), we can express as $\|T\mathbf{x}-T\mathbf{y}\|_{\mathcal{B}}$ as

$$\begin{aligned}
\|T\mathbf{x}-T\mathbf{y}\|_{\mathcal{B}} &= \left\|\Lambda\left[L^{-1}\Gamma_x(L) - L^{-1}\Gamma_y(L)\right]_+ \mathbf{w}_t\right\|_{\mathcal{B}} && \text{(C.9)}\\
&\leq \|\Lambda\|_{\mathcal{B}}\left\|\left[L^{-1}\Gamma_x(L) - L^{-1}\Gamma_y(L)\right]_+ \mathbf{w}_t\right\|_{\mathcal{B}} && \text{(C.10)}\\
&\leq \|\Lambda\|_{\mathcal{B}}\left\|\left[\Gamma_x(L) - \Gamma_y(L)\right]\mathbf{w}_t\right\|_{\mathcal{B}} && \text{(C.11)}\\
&\leq \alpha\|\mathbf{x}-\mathbf{y}\|_{\mathcal{B}}. && \text{(C.12)}
\end{aligned}$$

The inequality on the second line follows from the definition of the operator norm (6.8). The inequality on the third line follows from that

$$\begin{aligned}
\left\|\left[L^{-1}\Gamma_x(L) - L^{-1}\Gamma_y(L)\right]_+ \mathbf{w}_t\right\|_{\mathcal{B}} &= \sum_{i=1}^{n}\sqrt{\sum_{\tau=1}^{m}\sum_{s=1}^{\infty}\left(\gamma_{i,\tau,x,s} - \gamma_{2,\tau,y,s}\right)^2} && \text{(C.13)}\\
&\leq \sum_{i=1}^{n}\sqrt{\sum_{\tau=1}^{m}\sum_{s=0}^{\infty}\left(\gamma_{i,\tau,x,s} - \gamma_{2,\tau,y,s}\right)^2} && \text{(C.14)}\\
&= \|\mathbf{x}-\mathbf{y}\|_{\mathcal{B}}. && \text{(C.15)}
\end{aligned}$$

The inequality on the last line (C.12) follows from the condition (C.3) on the operator norm of Λ. □

That a system of the form (6.1)-(6.4) has a unique solution then follows from the contractive property of T.

Appendix D. Exogenous signals

The proofs of Proposition 1 and Theorem 1 used that agents can observe the endogenous variables in $\mathbf{p}_t$. Here we show that the mapping T is a contraction also when agents only observe the vector $\mathbf{z}_{t,j}$ of exogenous signals. Comparing the distances $\|\mathbf{x}-\mathbf{y}\|_{\mathcal{B}}$ and $\|T\mathbf{x}-T\mathbf{y}\|_{\mathcal{B}}$ for $\mathbf{x},\mathbf{y}\in\mathcal{B}^n$ is then substantially simpler.

Proposition 2. *(T is a contraction with exogenous signals) Define the mapping* $T:\mathcal{B}^n\rightarrow\mathcal{B}^n$

$$T\mathbf{x}=\Lambda\int E\left[\mathbf{x}_{t+1}\mid\Omega_{t,j}^{z}\right]dj+F_{\Theta}\mathbf{\Theta}_t+F_w\mathbf{w}_t \tag{D.1}$$

for $\mathbf{x}\in\mathcal{B}^n$ *and where* $\Omega_{t,j}^{z}\equiv\left\{\mathbf{z}_{t,j},\Omega_{t-1,j}\right\}$. *T is then a contraction mapping if* $|eig(M_0)|<1$ *and* $\|\Lambda\|_{\mathcal{B}}<\alpha$.

Proof. That T is a self-map on $\mathcal{B}^n$ when $|eig(M_0)|<1$ follows from Lemma 3. Define the average agent in period t as the agent whose expectations coincide with the cross-sectional average expectation. Denote the the projection onto the space spanned by the history of exogenous signals observed by the average agent in period t as $\mathcal{P}_z$. We can then express the distance $\|T\mathbf{x}-T\mathbf{y}\|_{\mathcal{B}}$ as

$$\begin{aligned}
\|T\mathbf{x}-T\mathbf{y}\|_{\mathcal{B}} &= \|\Lambda\mathcal{P}_z\mathbf{x}_{t+1}-\Lambda\mathcal{P}_z\mathbf{y}_{t+1}\|_{\mathcal{B}} && \text{(D.2)}\\
&= \|\Lambda\mathcal{P}_z\left(\mathbf{x}_{t+1}-\mathbf{y}_{t+1}\right)\|_{\mathcal{B}} && \text{(D.3)}\\
&\leq \|\Lambda\|_{\mathcal{B}}\|\mathcal{P}_z\left(\mathbf{x}_{t+1}-\mathbf{y}_{t+1}\right)\|_{\mathcal{B}} && \text{(D.4)}\\
&\leq \|\Lambda\|_{\mathcal{B}}\|\mathbf{x}-\mathbf{y}\|_{\mathcal{B}} && \text{(D.5)}\\
&< \alpha\|\mathbf{x}-\mathbf{y}\|_{\mathcal{B}}. && \text{(D.6)}
\end{aligned}$$

The equality on the second line follows from that the fact that projections are linear operators. The inequality on the third line follows from the definition of the operator norm (6.8). The inequality on the fourth line follows from the fact that the norm of a projection operator is 1, i.e. $\|\mathcal{P}_z\|_{\mathcal{B}}=1$ and from that (with a slight abuse of notation) $\|\mathbf{x}_{t+1}-\mathbf{y}_{t+1}\|=\|\mathbf{x}-\mathbf{y}\|$. The inequality on the last line, which completes the proof, follows from the condition that $\|\Lambda\|_{\mathcal{B}}<\alpha$. □

Appendix E. A Recursive Solution Algorithm for the General Model

To show that the equilibrium can be approximated to an arbitrary precision with a finite state, we need to modify the algorithm of Section 3 to allow for the more general form of (6.1)-(6.4).

Step 1. As a starting point, take the zero order process for $\mathbf{p}$

$$\mathbf{p}_t^0 = G_0\mathbf{\Theta}_t + F_w\mathbf{w}_t \tag{E.1}$$

$$\mathbf{\Theta}_t = M_0\mathbf{\Theta}_{t-1} + N_0\mathbf{w}_t \tag{E.2}$$

where

$$G_0 = F_\Theta \tag{E.3}$$

and the associated measurement equation

$$\mathbf{s}_{t,j}^0 = D_0\Theta_t^{(0:k)} + R_w\mathbf{w}_t + R_\eta\mathbf{w}_{t,j} : \mathbf{w}_{t,j} \sim N(0, I) \tag{E.4}$$

where

$$D_0 = \begin{bmatrix} D_\Theta \\ G_0 \end{bmatrix}, \quad R_w = \begin{bmatrix} R_{zw} \\ F_w \end{bmatrix}, \quad R_\eta = \begin{bmatrix} R_{z\eta} \\ \mathbf{0} \end{bmatrix}. \tag{E.5}$$

Step 2. Compute the matrices M_{k+1} and N_{k+1} as

$$M_{k+1} = \begin{bmatrix} M_0 & \mathbf{0}_{q\times kq} \\ \mathbf{0}_{kq\times q} & \mathbf{0}_{kq\times kq} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{q\times kq} & \mathbf{0}_{q\times q} \\ K_kD_kM_k & \mathbf{0}_{kq\times q} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{q\times q} & \mathbf{0}_{q\times kq} \\ \mathbf{0}_{kq\times q} & (I - K_kD_k)M_k \end{bmatrix} \tag{E.6}$$

$$N_{k+1} = \begin{bmatrix} N_0 \\ (K_kD_kN_k + K_kR_w) \end{bmatrix}. \tag{E.7}$$

to get the k^{th} step law of motion

$$\mathbf{\Theta}_t^{(0:k)} = M_k\Theta_{t-1}^{(0:k)} + N_k\mathbf{w}_t : \mathbf{w}_t \sim N(0, I) \tag{E.8}$$

where the matrix D_k defined as

$$D_k = \begin{bmatrix} D_\Theta \quad \mathbf{0}_{q\times kq} \\ G_k \end{bmatrix} \tag{E.9}$$

and K_k is the Kalman gain (A.1).

Step 3. Compute the k-order process $\mathbf{p}_t^{k+1}$ by using the vector and matrix equivalent of (3.30)

$$\mathbf{p}_t^{k+1} = G_{k+1}\Theta_t^{(0:k+1)} + F_w\mathbf{w}_t \tag{E.10}$$

where

$$G_{k+1} = F_\Theta + \Lambda G_kM_kH_{k+1} \tag{E.11}$$

and

$$H_k \equiv \begin{bmatrix} \mathbf{0}_{(kq)\times q} & I_{kq} \end{bmatrix}. \tag{E.12}$$

Step 4. Repeat Steps $2-3$ for $k = 1, 2, 3, ..., \overline{k}$.

The number of iterations $\overline{k}$ can be chosen to achieve to achive any desired degree of accuracy, using the approximation error bounds (4.18) and (4.19) but with α replacing β and the norm $\|\cdot\|_{\mathcal{B}}$ replacing $\|\cdot\|$.