Peering into the mist: social learning over an opaque observation network*

John Barrdear†
Bank of England and the Centre for Macroeconomics

4th June 2014

Abstract

I present a model of social learning over an exogenous, directed network that may be readily nested within broader macroeconomic models with dispersed information and combines the attributes that agents (a) act repeatedly and simultaneously; (b) are Bayes-rational; and (c) have strategic interaction in their decision rules. To overcome the challenges imposed by these requirements, I suppose that the network is *opaque*: agents do not know the full structure of the network, but do know the link distribution. I derive a specific law of motion for the hierarchy of aggregate expectations, which includes a role for *network shocks* (weighted sums of agents' idiosyncratic shocks). The network causes agents' beliefs to exhibit increased persistence, so that average expectations overshoot the truth following an aggregate shock. When the network is sufficiently (and plausibly) irregular, transitory idiosyncratic shocks cause persistent aggregate effects, even when agents are identically sized and do not trade.

JEL classification: C72, D82, D83, D84
Key words: dispersed information, network learning, heterogeneous agents, aggregate volatility

*The views expressed in this paper are those of the author and not necessarily those of the Bank of England. This paper was developed at the London School of Economics before the author's employment by the Bank of England. All errors are my own.

†I am indebted to Kevin Sheedy, Wouter DenHaan, Kristoffer Nimark, Andrea Prat, Francesco Nava, Tony Yates, an anonymous referee and various seminar and workshop participants at the London School of Economics and the Bank of England for their invaluable feedback. This work was partially funded by the ESRC.
Email: john.barrdear@bankofengland.co.uk

1 Introduction

This paper attempts a partial bridging of three strands of research, namely those of *network learning* in the microeconomic theory literature; *dispersed information* in macroeconomic modelling; and more recent work identifying *idiosyncratic origins of aggregate volatility.* Although models at this intersection are not generally solvable, with a plausible restriction on the observability of the network structure I am able to derive a law of motion for the full hierarchy of expectations and demonstrate that a finite approximation of the same may be made arbitrarily accurate. I further show that network learning causes aggregate beliefs to be more persistent than the shocks that cause them, so that average expectations *overshoot* the truth. The observation network also provides a channel for idiosyncratic shocks to cause aggregate volatility even when all agents are identically sized and do not trade with each other.

The *dispersed information* literature, which started with Woodford (2003), is one of three strands of research (the others being *sticky information* and *rational inattention*[1]) that seek to reintroduce the ideas of Lucas (1972) and Phelps (1984) – that information frictions are crucial to explaining the dynamics of aggregate variables following a shock.

In particular, Woodford invoked the central insight of Townsend (1983) – that with heterogenous information and strategic interaction, rational agents become interested in an infinite regress of higher-order beliefs – and demonstrated that because of the sluggish response of higher-order expectations, aggregate rigidity broadly equivalent to that produced by Calvo (1983) pricing may be replicated in a model with fully flexible price-setting firms observing independent and unbiased signals of nominal GDP. A flurry of further work has ensued,[2] but three key attributes are typically seen as essential features of such models, beyond the simple fact of a hidden dynamic state and heterogeneous information:

1. agents act repeatedly;
2. agents update their beliefs in a Bayesian and model-consistent (i.e. rational) manner;[3] and
3. agents act strategically, with their payoffs a function of other players' actions.

These requirements are standard and largely uncontroversial features of macroeconomic modelling. The first is a defining feature of any dynamic model, the second is considered necessary to address the Lucas (1976) critique and the third is both analytically necessary to generate the higher-order expectations of Townsend (1983) and widely observed in a variety of financial and macroeconomic settings, including firms' price-setting; search and matching models; and financial asset pricing.

The addition of *network learning* – whereby individual agents observe the actions of specific competitors in order to learn about a hidden state – would appear a natural extension to the dispersed information literature, particularly when acquiring comprehensive information would be prohibitively costly. For example, a firm that experiences shocks to its demand or its marginal costs but does not

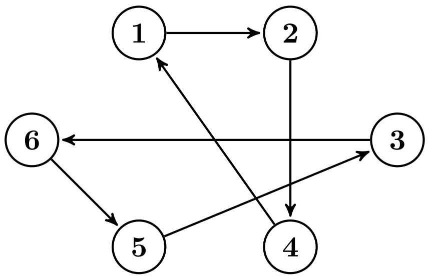
[1]The sticky information literature derives from Mankiw and Reis (2002, 2006, 2007), while the rational inattention literature dates to Sims (2003).

[2]See, for example, Nimark (2008), Lorenzoni (2009), Angeletos and La'O (2009, 2010), Graham and Wright (2010), Graham (2011a,b) and Melosi (2014).

[3]The *learning* literature (see, e.g., Evans and Honkapohja, 2001) explores settings under which non-model-consistent expectations converge to model-consistent expectations. A model by Graham (2011b) demonstrates that such dynamics at the aggregate level are dominated by dispersed information (i.e. that learning converges quickly to model-consistent expectations relative to the persistence of dispersed expectation errors).

know whether these are common to all firms can partially infer the average by observing the price of a competitor. A trader wanting to learn the level of demand for a given asset can improve their standing by speaking to other traders. Households that experience complementarity in their consumption will improve their welfare by observing the choices of their neighbours.

A model of network learning that possesses all of the above attributes is notoriously difficult to study, however. Solving such a model, let alone simulating it or nesting it within a broader model of the economy, has typically been thought to be sufficiently great as to preclude comprehensive analysis in anything other than trivially small networks (Jackson, 2008). As such, the literature to date has proceeded by avoiding one or more of the above assumptions (see below for a brief review).



(a) Graph representation

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

(b) Adjacency matrix

Figure 1: An example of an observation network. Arrows indicate the direction of observation (agent 1 observes agent 2, etc). It is "regular" in that all agents observe, and are observed by, the same number of competitors. The adjacency matrix (G) is such that $G_{ij} = 1$ if node i observes node j and zero otherwise.

In contrast to earlier work, the present paper is able to embrace all three of these assumptions by combining them with a fourth: network opacity. By denying agents knowledge of the exact topology of the network (the network is *opaque*) and instead supposing that they know only the distribution from which observation targets are drawn, I derive the law of motion for the full hierarchy of agents' expectations and show that *network shocks* (weighted sums of agents' idiosyncratic shocks) enter at an aggregate level.[4] The researcher is therefore able to simulate the aggregate effects of network learning without a need to simulate individual agents' decisions. This makes the model particularly amenable to nesting within broad general equilibrium models of the economy.

With an opaque network, agents switch from considering their competitors' *individual* beliefs to instead contemplating a sequence of *weighted averages* of all agents' beliefs. Since agents' learning is recursive, this allows the curse of dimensionality to be be overcome in practice, as an arbitrarily accurate approximation of the full solution can be found by selecting a sufficiently high cut-off for the number of weighted-averages to include, together with the standard cut-off for the number of higher-orders of expectation.

The imposition of an opaque network is both intuitive and appealing. It is not plausible, for example, to suppose that every business knows to whom every other business speaks, just as nobody knows the identity of all of their friends' friends. From the researcher's perspective, this ignorance of topology makes it particularly challenging when attempting to consider the aggregate effects of network learning. But by recognising that not only the researcher but also the economic agents themselves are ignorant of the network structure, the researcher can identify laws of motion for the agents' aggregate beliefs, even if they can never pin down the path of any individual's expectation.

[4]With the underlying (and unobservable) state following an AR(1) process, the full hierarchy of expectations about it will follow an ARMA(1,1) process, with network shocks entering both contemporaneously and with a lag.

The second requirement – that agents not learn about the structure of the network over time – may be thought of in two ways. First, one might consider a setting in which the network is dynamic, changing every period. *In extremis*, this would involve the network being destroyed and redrawn each period, so that agents are not *able* to learn about the network as it does not persist over time. Alternatively, one might suppose that the network was drawn once, at time zero, but agents are boundedly rational in that they do not attempt to learn about it beyond the common knowledge of the distribution from which it was drawn. In this setting, agents' decisions are perhaps best described as *conditionally rational*, in that conditional on the structure of the network, they are rational in their processing of the information they gain from it.

For sufficiently irregular networks – i.e. where some agents' actions are disproportionately observable – I show that network shocks do not converge to zero and therefore add aggregate volatility to the system, even when all agents are the same size (e.g. even when all firms contribute the same share of aggregate production). Despite idiosyncratic shocks being purely transitory, the aggregate volatility they induce through the network is also shown to exhibit (endogenous) persistence.

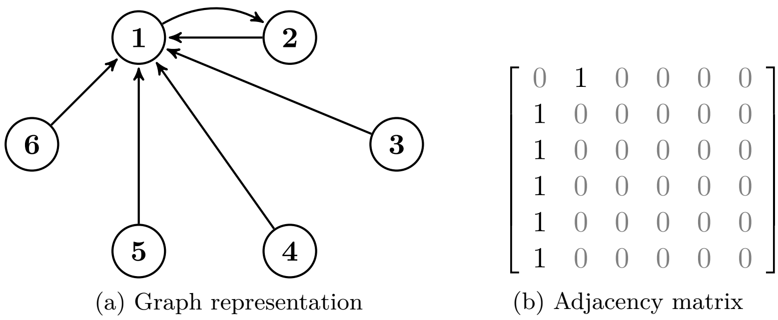


Figure 2: An irregular observation network. Agent 1 observes agent 2, while all others observe agent 1.

This paper therefore adds to the burgeoning literature on deriving aggregate volatility from agents' idiosyncratic shocks, within which the most closely related work is that of Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Saleh (2012), who examine a static model of firms operating within an inter-sectoral supply network (focussing on the cumulative effect of the network as a transmission mechanism of real shocks). By contrast, I present a dynamic model that considers the evolution of higher-order beliefs over each round of learning. Acemoglu *et al*'s emphasis on what they call "higher-order interconnectivity" in the network is captured and given an explicit dynamic role here. Finally, the observation network explored here may also clearly be different to the trading network of an economy.

In another vein, Gabaix (2011) demonstrates how aggregate volatility can emerge from idiosyncratic shocks when the distribution of firm sizes exhibits fat tails, even when those firms do not trade directly with each other. Each of these share with the current paper an emphasis on unequal, or fat-tailed, distributions. In the model of Gabaix (2011), aggregate volatility arises because the largest firms contribute disproportionately to aggregate production, while in that by Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Saleh (2012) it emerges through those firms whose output is most extensively used as an intermediate good by other firms. By contrast, the current paper demonstrates a granular origin of aggregate volatility even for identically sized agents that do not trade with each other by focussing, instead, on firms' signal extraction problem.

The model generates average expectations that are more persistent than the shocks that cause them, so that following an innovation to the underlying state, impulse responses overshoot the underlying state in a cascade effect that combines the herding exhibited in both Banerjee (1992), where agents observe others' actions, but have no strategic motive; and Morris and Shin (2002), where agents have a strategic motive, but do not observe each others' actions. The degree of persistence is shown to be increasing in the number of agents observed.

The intuition for this is as follows. Since the network is opaque, agents cannot know who their observee is watching. Common knowledge of the distribution from which observees are drawn, together with the linearity of the model, means that all agents treat all other agents *as though* they observe the same weighted average of everybody's action. In a model with strategic complementarity and dispersed information, public signals represent a source of herding, as famously shown by Morris and Shin (2002). With network learning over an opaque network, the observation of a competitor's action is therefore a signal of a hypothetical public signal that, to the best of each agent's knowledge, everyone else effectively observes. Bayesian updating therefore causes them to place extra weight on observations of other agents' actions.

Methodologically, this paper expands on the work of Nimark (2008, 2011a,b), who in turn extended that of Woodford (2003). While Woodford only granted agents signals of the underlying state, Nimark also permitted agents to observe, with a lag, aggregate variables that depend on the entire hierarchy of expectations. This addition required the development of a new solution methodology that I here extend to the idea of agents observing the previous-period actions of specific competitors.

Although the models of Woodford and Nimark focus on firms' price-setting behaviour, the model developed here is context free and may be applied to any general setting with strategic interaction and network learning. The conclusion considers a number of examples of such applications.

The remainder of this paper is organised as follows. The remainder of this introduction first provides a brief survey of previous models of network learning. Section 2 then provides some preliminary definitions related to graph theory, hierarchies of expectations and asymptotically non-uniform distributions. Section 3 next presents the general model, together with a characterisation of the solution and a methodology for finding it. Section 4 provides an illustrative example of the model in action, applying it to the commonly-used decision rule examined by Morris and Shin (2002) and Calvó-Armengol and de Martí (2007). Section 5 concludes.

1.1 Existing literature on network learning[5]

As mentioned above, literature on network learning has, to date, proceeded by avoiding one or more of the three assumptions that (a) agents are rational; (b) agents act simultaneously and repeatedly over many periods; and (c) agents' optimal decisions include consideration of strategic complementarity. Early work in observational learning, for example, focussed on *sequential learning*, with each agent making a single, irrevocable decision in an exogenously defined order, typically after observing the actions of all, or a well-defined subset, of their predecessors. In such a setting, it is well known that agents can rationally (in the Bayesian sense) exhibit "herding", or "information cascades", whereby their private signals regarding the unknown state are swamped by the weight of past actions (see, for example, Banerjee, 1992; Lee, 1993; and Smith and Sørensen, 2000).

More recently, work in sequential learning has examined situations where the observation neighbourhood of each agent is determined stochastically. Banerjee and Fudenberg (2004), for example,

[5] Acemoglu and Ozdaglar (2011) also provide a recent review.

demonstrate that convergence will occur if the sampling of earlier players' beliefs is "unbiased" in the sense that it is representative of the population as a whole and at least two earlier players are sampled. More generally, Acemoglu, Dahleh, Lobel, and Ozdaglar (2011) characterise the (Bayesian) equilibrium of a sequential learning model for a general stochastic sampling process. They demonstrate that so long as no group of agents is excessively influential, there will be asymptotic learning of the truth when private beliefs are unbounded[6] and characterise some settings under which asymptotic learning still emerges when private beliefs are bounded.

Although this more recent work carries the flavour of network learning in that agents observe the actions of only a subset of their competitors,[7] they do not meet the popular conception of network learning in which agents undertake repeated, simultaneous actions in an environment of strategic interaction. Tackling such a problem, however, is notoriously difficult. The presence of strategic interaction introduces the need to consider the infinite hierarchy of higher-order (average) beliefs. When agents exist in an observation network, it is also necessary for each of them to consider the specific belief held by their observation target and, in turn, the belief of their target's target and so forth. As the number of agents in the network expands, this causes an explosion in the size of the state vector quite apart from the presence of higher-order expectations (see section 2.2 below for more detail), thereby subjecting the problem to the famous curse of dimensionality.

In order to analyse learning in a repeated, simultaneous action environment, the literature has therefore most commonly chosen to abandon the assumption of Bayesian updating. Non-Bayesian learning over a network is typically modelled in the style of DeGroot (1974), with agents applying a constant weight to their observations of competitors' actions. For example, DeMarzo, Vayanos, and Zwiebel (2003) explore situations where agents assume that signals they receive from observing each other contain *entirely new information*. In a setting where a finite number of agents wish to estimate an unknown, but fixed state $\theta \in \mathbb{R}^L$, they suppose that agents each receive a single, conditionally independent and unbiased signal of the state and then communicate their beliefs over multiple rounds. Imposing the assumption that agents update their beliefs via a simple and constant weighted sum greatly simplifies analysis, but introduces what the authors label "persuasion bias" from the agents' failure to properly discount the repetition of information they receive. Calvó-Armengol and de Martí (2007) extend this setting to provide an assessment of the welfare losses from "unbalanced," or irregular[8] networks.

Golub and Jackson (2010) likewise study learning in a setting where agents "naïvely" update their beliefs regarding a fixed state of the world by taking weighted averages of their neighbour's opinions. In contrast to earlier work, they are able demonstrate that with such heuristic learning, individual beliefs converge to the truth for a broad variety of networks (provided they are sufficiently large) and provide upper and lower bounds on the rate of convergence.

In the area of what might be called "true" Bayesian network learning (repeated simultaneous actions with agents engaged in Bayesian updating), there has been remarkably little work to date. Gale and Kariv (2003) examine Bayesian network learning in a setting with *connected* networks[9] and in which agents' payoffs depend only on the proximity of their expectation to the state (i.e. without

[6]That is, where agents may receive arbitrarily strong signals so that the support of their posterior belief that the state is equal to a given possibility is $[0, 1]$.

[7]Indeed, Acemoglu, Dahleh, Lobel, and Ozdaglar (2011) refer to their model as one of learning over a social network.

[8]A regular network is one in which all nodes have the same number of inbound and outbound links.

[9]In this context, a connected network is one in which information is able to flow *from* any agent *to* any other agent.

any strategic interaction). They note that the "computational difficulty of solving the model is massive even in the case of three persons." Mueller-Frank (2013) details a formal structure for Bayesian learning over an undirected social network (i.e. with pairwise sharing), allowing for a choice correspondence from information to actions (and general strategies for the selection between indifferent options) as opposed to outright decision rules, but notes the extreme practical difficulties of actually implementing such a rule, both for the agents in principle and the researcher more generally.

Furthermore, both Gale and Kariv (2003) and Mueller-Frank (2013) step away from consideration of strategic interaction in agents' decision-making, so that when observing any competitor, every agent knows that their action is driven entirely by their belief regarding the underlying state.

2 Preliminary definitions

2.1 Network terminology

Only terms necessary for this paper are provided here. Readers interested in a more comprehensive introduction are directed to more general texts on networks in economics.[10]

A *network*, or *graph*, is a collection of nodes and edges.[11] In this paper, each agent will be a node in a network. A network's *adjacency matrix*, G, is such that $G_{ij} = 1$ if node i is connected to node j and zero otherwise. An *undirected* network is one in which $G_{ij} = G_{ji}\ \forall i, j$, while a *directed* network permits $G_{ij} \neq G_{ji}$. Figure 1 illustrates an example of a directed network comprised of six nodes.

For a directed network, the *out-degree* of a node is the number of edges originating at that node ($d_i^{out} = \sum_{j=1}^{n} G_{ij}$), while the *in-degree* of a node is the number of edges arriving at it ($d_j^{in} = \sum_{i=1}^{n} G_{ij}$).

A *regular* network is one in which all nodes have exactly the same out-degree and in-degree, so that $d_i^{out} = d_i^{in} = \overline{d}\ \forall i$. An *irregular* network is one in which this is not the case. The network shown in figure 1 is regular, while that shown in figure 2 is irregular.

In the model of this paper, all nodes will have the same out-degree in that every agent will observe the same number of other agents ($d_i^{out} = q\ \forall i$). I will therefore interchangeably refer to agents' in-degree as simply their *degree* (d_i). The *degree sequence* of a network is the set $\{d_1, d_2, \cdots, d_n\}$. Without loss of generality, I assume that nodes (agents) are arranged such that $d_1 \geq d_2 \geq \cdots \geq d_n$.

A network is *connected* if it is undirected and a path exists from any node to any other node. A network is *strongly connected* if it is directed and a route (a directed path) exists from any node to any other node. A network is *complete* if every node is directly connected to every other node so that all possible edges exist.

A *cycle* of length k is a sequence of nodes, starting and ending with the same node, $\{i_1, \cdots, i_{k-1}, i_k, i_1\}$ such that each pair of consecutive nodes are connected ($G_{i_j i_{j+1}} = 1$). A network is *aperiodic* if the greatest common divisor of the lengths of its cycles is one. In other words, a network is aperiodic if at least one node has a link to itself ($\exists i : G_{ii} = 1$).

In an environment of social learning where agents share their beliefs truthfully, a network being aperiodic implies that at least one agent places non-zero weight on their own prior when updating their belief following the observation of their neighbours. In such a setting, if the network is both

[10] See, for example, Goyal (2007) or Jackson (2008).

[11] Some papers refer to a network as a *weighted graph* – i.e. a graph with a weight associated with each edge.

aperiodic and strongly connected, it is well known that all agents' beliefs will converge to the same values (Kemeny and Snell, 1960).

In the current paper, all agents will receive private, unbiased signals in every period. With Bayesian updating, every *sub-graph* of the network will therefore be aperiodic, so that convergence is assured and uninteresting. Instead, this paper focusses on the dynamics of aggregate beliefs over time.

2.2 Higher-order expectations

The near-ubiquitous treatment of higher-order expectations in economic literature to date[12] has considered only the hierarchy of *simple average* expectations. That is, to consider settings where agents are interested only in the sequence of objects $\left\{x_t, \overline{E}_t\left[x_t\right], \overline{E}_t\left[\overline{E}_t\left[x_t\right]\right], \cdots\right\}$ where $\overline{E}_t\left[\cdot\right] \equiv \int_0^1 E_t\left(i\right)\left[\cdot\right] di$.[13]

This is a modelling choice only, however, made for analytical convenience. In particular, it is not appropriate for a model of learning over a network where economic agents must (in principle, at least) form opinions regarding the beliefs of every other agent in the network and know that they will each, in turn, do the same. To model these fully, it is necessary to work with a more generalised definition of a hierarchy of expectations.

Definition 1. *A* ***compound expectation*** *is a weighted average of all agents' expectations.*

For example, let $\boldsymbol{x}_t$ be an $(m \times 1)$ vector of random variables, $E\left[\boldsymbol{x}_t | \mathcal{I}_t\left(i\right)\right]$ be the expectation of $\boldsymbol{x}_t$ conditioned on the period t information set of agent i and let $\boldsymbol{w}$ be an $(n \times 1)$ vector of weights such that $w_i \in [0,1]$ and $\sum_{i=1}^n w_i = 1$. The compound expectation $E_{\boldsymbol{w},t}\left[\boldsymbol{x}_t\right]$ is given by:

$$E_{\boldsymbol{w},t}\left[\boldsymbol{x}_t\right] \equiv \begin{bmatrix} E\left[\boldsymbol{x}_t|\mathcal{I}_t\left(1\right)\right] & E\left[\boldsymbol{x}_t|\mathcal{I}_t\left(2\right)\right] & \cdots & E\left[\boldsymbol{x}_t|\mathcal{I}_t\left(n\right)\right] \end{bmatrix} \boldsymbol{w} = \sum_{i=1}^{n} w_i E_t\left[\boldsymbol{x}_t|\mathcal{I}_t\left(i\right)\right] \tag{1}$$

Note that this nests both simple, or unweighted, average expectations (e.g. $\boldsymbol{w}_A = \begin{bmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix}'$) and individual expectations (e.g. $\boldsymbol{w}_B = \begin{bmatrix} \mathbf{0}' & 1 & \mathbf{0}' \end{bmatrix}'$). With the usual notation that the 0^{th}-order expectation of a variable is the variable itself, we next define:

Definition 2. *A* ***hierarchy of expectations****, from order* 0 *to* k*, is defined* recursively *as:*

$$\mathbb{E}_t^{(0:k)}\left[\boldsymbol{x}_t\right] = \begin{bmatrix} \boldsymbol{x}_t \\ E_{\boldsymbol{w}_A,t}\left[\mathbb{E}_t^{(0:k-1)}\left[\boldsymbol{x}_t\right]\right] \\ E_{\boldsymbol{w}_B,t}\left[\mathbb{E}_t^{(0:k-1)}\left[\boldsymbol{x}_t\right]\right] \\ \vdots \end{bmatrix} \tag{2}$$

This is not simply the stacking of each order of expectations on top of each other. For example, if $\boldsymbol{x}_t$ is scalar and there are two compound expectations, the hierarchies $(0:1)$ and $(0:2)$ are given by:

[12] While modern macroeconomic literature on higher-order expectations dates to Townsend (1983), the general idea has been known since, at least, the famous "beauty contest" argument of Keynes (1936).

[13] One recent exception is Kohlhas (2013), who examines the value of central bank disclosure in a model with two compound expectations – that of the central bank and the average expectation of the private sector. Kohlhas extends the solution methodology of Nimark (2008, 2011a) in a similar manner to the current paper.

$$\mathbb{E}_t^{(0:1)}[\boldsymbol{x}_t] = \begin{bmatrix} \boldsymbol{x}_t \\ E_{\boldsymbol{w}_A,t}[\boldsymbol{x}_t] \\ E_{\boldsymbol{w}_B,t}[\boldsymbol{x}_t] \end{bmatrix} \quad \mathbb{E}_t^{(0:2)}[\boldsymbol{x}_t] = \begin{bmatrix} \boldsymbol{x}_t \\ E_{\boldsymbol{w}_A,t}\begin{bmatrix} \boldsymbol{x}_t \\ E_{\boldsymbol{w}_A,t}[\boldsymbol{x}_t] \\ E_{\boldsymbol{w}_B,t}[\boldsymbol{x}_t] \end{bmatrix} \\ E_{\boldsymbol{w}_B,t}\begin{bmatrix} \boldsymbol{x}_t \\ E_{\boldsymbol{w}_A,t}[\boldsymbol{x}_t] \\ E_{\boldsymbol{w}_B,t}[\boldsymbol{x}_t] \end{bmatrix} \end{bmatrix}$$

The benefit of depicting hierarchies in this recursive manner is that it becomes both conceptually and computationally simple to extract sub-hierarchies comprised of a single compound expectation. For example, if $E_{\boldsymbol{w}_A,t}[\boldsymbol{x}_t] = \overline{E}_t[\boldsymbol{x}_t] = \overline{\boldsymbol{x}}_{t|t}^{(1)}$ is the simple-average expectation, then the sub-hierarchy of $\overline{\boldsymbol{x}}_t^{(0:k)}$ may be extracted as: $\overline{\boldsymbol{x}}_t^{(0:k)} = \begin{bmatrix} I & 0 \end{bmatrix} \mathbb{E}_t^{(0:k)}[\boldsymbol{x}_t]$. This recursive formulation of the expectation hierarchy is also a necessary feature of the solution methodology developed below.

Size of the expectation hierarchy

When $\boldsymbol{x}_t$ contains m elements and there are p compound expectations of interest, the set of k^{th}-order expectations will contain mp^k distinct elements. However, it does not generally follow that the hierarchy $\mathbb{E}_t^{(0:k^*)}[\boldsymbol{x}_t]$ will contain $m\left(\sum_{k=0}^{k^*} p^k\right)$ unique elements. This is because if one of the compound expectations, say $E_{\boldsymbol{w}_B}[\cdot]$, is formed from a single information set then the law of iterated expectations implies that $E_{\boldsymbol{w}_B,t}[E_{\boldsymbol{w}_B,t}[\boldsymbol{x}_t]] = E_{\boldsymbol{w}_B,t}[\boldsymbol{x}_t]$. In general, when $q\,(\leq p)$ is the number of individual expectations, the number of unique elements in the hierarchy $\mathbb{E}_t^{(0:k^*)}[\boldsymbol{x}_t]$ will be:[14]

$$N(m,p,q,k^*) = m\left(p^{k^*} + \sum_{k=0}^{k^*-1}\left(p^k - q\sum_{s=0}^{k} p^s\right)\right) \tag{3}$$

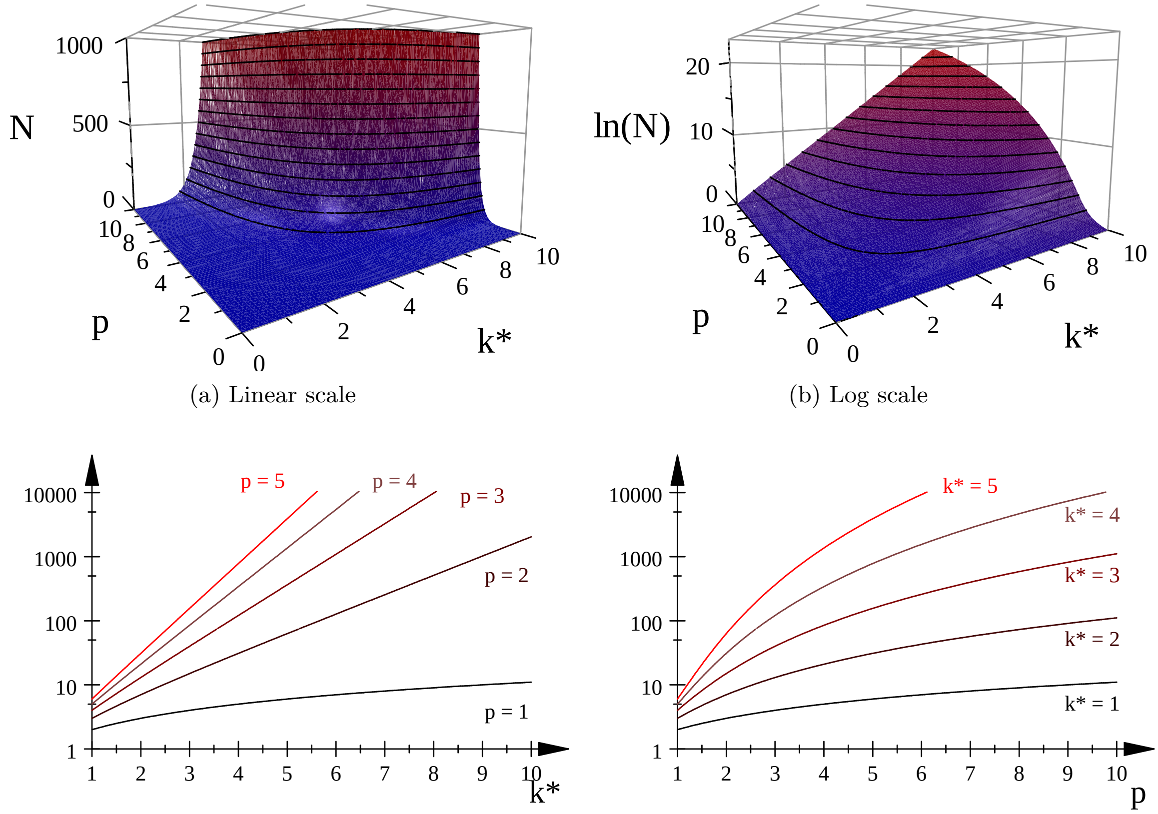
with $N(m,p,0,k^*) = m\left(\sum_{k=0}^{k^*} p^k\right)$. Even when $q = p$, though, it should be readily apparent that the size of an expectation hierarchy explodes in both p (the number of compound expectations) and k^* (the highest order in expectations), as figure 3 shows.

Approximations of model solutions must therefore be found by limiting attention to a finite subset of the full state. The vast majority of models in the dispersed information literature have $p = 1$ and place decreasing weight on higher-order expectations (i.e. the weight is decreasing in k). Provided that the variance of higher-order expectations remains bounded from above, these models can be approximated to an arbitrary degree of accuracy by imposing a limit, k^*, on the number of orders of expectation and including all orders from zero up to that cut-off.

In contrast, increasing the number of relevant compound expectations can be more problematic as there is rarely, if ever, an obvious reason for weighting them differently. In network learning, in particular, p will be equal to the number of nodes when the network is strongly connected.[15]

[14]
$$m\left(\underbrace{[1]}_{0^{th}\text{ order}} + \underbrace{[p]}_{1^{st}\text{ order}} + \underbrace{[p^2-q]}_{2^{nd}\text{ order}} + \underbrace{[p*(p^2-q)-q]}_{3^{rd}\text{ order}} + \underbrace{[p*(p*(p^2-q)-q)-q]}_{4^{th}\text{ order}} + \cdots\right)$$
$$= m\left(\left(\sum_{k=0}^{k^*} p^k\right) - q\left(\sum_{k=0}^{k^*-1}\sum_{s=0}^{k} p^s\right)\right), \text{ which rearranges to the equation in the text}$$

[15] Since in a strongly connected observation network information will flow from all nodes to all nodes, the construction of agents' priors requires that they consider the beliefs of all other agents.



(a) Linear scale (b) Log scale

(c) Cross-section by p (log scale) (d) Cross-section by k^* (log scale)

Figure 3: The number of elements in an expectation hierarchy ($q = 0$, $m = 1$)

2.3 Asymptotically non-uniform distributions

This paper develops a novel channel through which idiosyncratic shocks need not "wash out" and may, instead, induce aggregate volatility. This emerges because of agents' need to consider *weighted* averages of other agents' idiosyncratic shocks. In particular, a sufficient condition for such weighted sums to not converge to zero as $n \to \infty$ is to suppose that the weights are asymptotically non-uniform:

Definition 3. *Let Φ_n be a discrete distribution with corresponding p.d.f.*[16] *$\phi_n(i)$ and let $\zeta_n \equiv \sum_{i=1}^{n} \phi_n(i)^2$ be the Herfindahl index of the same. Φ_n is **asymptotically non-uniform** if:*

- $\lim_{n\to\infty} \phi_n(i) = 0 \ \forall i$*; and*
- $\lim_{n\to\infty} \zeta_n = \zeta^*$ *where* $\zeta^* \in (0, 1)$.

To illustrate the emergence of aggregate volatility, suppose that each agent receives an independent, mean zero shock drawn from a common Gaussian distribution – $\boldsymbol{v}(i) \sim N(\mathbf{0}, \Sigma_{vv}) \ \forall i$ – and consider a setting where a weighted average of agents' shocks is economically significant:

$$\tilde{\boldsymbol{v}}_n \equiv \sum_{i=1}^{n} \boldsymbol{v}(i)\, \phi_n(i) \text{ where } \phi_n(i) \in (0, 1) \text{ and } \sum_{i=1}^{n} \phi_n(i) = 1$$

[16] Strictly, for a discrete distribution, it is a probability *mass* function. But since this paper is concerned only with the limiting case of $n \to \infty$ assumes that they are indexed uniformly from zero to one so that the distribution becomes continuous, I stick with the conventional nomenclature.

Since $\tilde{\boldsymbol{v}}_n$ is a linear combination of mean zero Gaussian variables, it must itself be Gaussian with a mean of zero. Its variance will then be given by:

$$Var\left[\tilde{\boldsymbol{v}}_n\right] = \sum_{i=1}^{n} Var\left[\boldsymbol{v}\left(i\right)\phi_n\left(i\right)\right] di = \sum_{i=1}^{n} \Sigma_{vv}\phi_n\left(i\right)^2 di = \zeta_n\Sigma_{vv}$$

where the first equality relies on the shocks' independence to ignore covariance terms. The limiting variance as $n \to \infty$ is $\zeta^*\Sigma_{vv}$ and, hence, so long as $\zeta^* \neq 0$, the law of large numbers does not apply.

The set of asymptotically non-uniform distributions is quite broad, but in particular it includes the discrete power law distribution:

$$\phi_n\left(i\right) = c_n i^{-\gamma}; \text{ where } c_n = \left(\sum_{i=1}^{n} i^{-\gamma}\right)^{-1} \text{ and } \gamma > 1$$

and its equivalent for infinite n, the Zeta distribution. The shape parameter, $\gamma > 1$, governs the scaling of the distribution's tail, with larger values of γ corresponding to greater non-uniformity. Figure 4 plots the values of ζ^* for a range of values of γ for the Zeta distribution.[17]

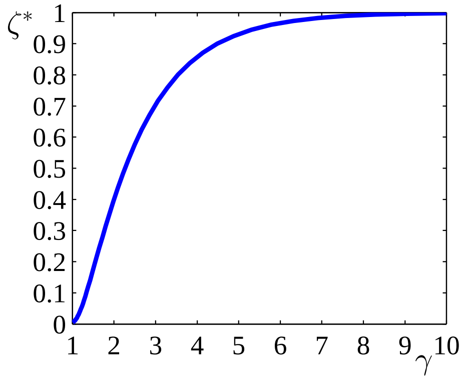


Figure 4: A plot of ζ^* for power law distributions with shape parameter γ

A great many observed networks, from links between pages on Wikipedia to established relationships in social networks, have been shown to be well approximated by power law degree sequences (i.e. the networks are *scale free*). See, for example, the work of Albert and Barabasi (2002), Jackson and Rogers (2007) or Clauset, Shalizi, and Newman (2009). It is important to appreciate, though, that I do not generally assume any particular distribution, only that it remains non-uniform (in the sense of definition 3) as the support of that distribution grows arbitrarily large.

3 The Model[18]

3.1 The general setting

There is a countably infinite number of agents,[19] indexed in a continuum between zero and unity.[20] The *underlying state* follows a vector autoregressive process:

$$\boldsymbol{x}_t = A\boldsymbol{x}_{t-1} + P\boldsymbol{u}_t \tag{4}$$

[17] Strictly, these are calculated for Zipf distributions with $n = 10^8$.

[18] Unless otherwise indicated, all proofs are provided in the appendix.

[19] An infinite number of agents is assumed to allow an appeal to relevant laws of large numbers when considering simple averages of zero-mean shocks.

[20] The assumption of indexing agents from zero to one is innocuous and made only to simplify the calculation of averages (e.g. $\overline{g}_t = \int_0^1 g_t\left(i\right) di$).

where $\boldsymbol{u}_t$ is a vector of shocks with mean zero, while A and P are fixed and publicly known. Agents do not observe the value of $\boldsymbol{x}_t$ and must instead form beliefs about it. I define X_t as the hierarchy of expectations regarding $\boldsymbol{x}_t$, in the sense of definition 2, and refer to it as the *state vector of interest.*

$$X_t \equiv \mathbb{E}_t^{(0:\infty)} [\boldsymbol{x}_t] \tag{5}$$

At a minimum, X_t contains $\boldsymbol{x}_t$ and the hierarchy of at least one compound expectation. For illustrative purposes, I will assume that agents' primary concern is with the the hierarchy of *simple-average* expectations, so that

$$\overline{\boldsymbol{x}}_{t|t}^{(0:\infty)} \in X_t \quad \text{where} \quad \overline{\boldsymbol{x}}_{t|t}^{(0:\infty)} \equiv \left[\boldsymbol{x}_t' \quad \overline{E}_t [\boldsymbol{x}_t]' \quad \overline{E}_t \left[\overline{E}_t [\boldsymbol{x}_t]\right]' \cdots\right]'$$

but it will be shown below that X_t must also include a variety of other compound expectations.

Agents' decision rule

Agents' actions are determined simultaneously and according to a common linear decision rule:

$$g_t (i) = \boldsymbol{\lambda}_1' E_t (i) [X_t] + \boldsymbol{\lambda}_2' \boldsymbol{x}_t + \boldsymbol{\lambda}_3' \boldsymbol{v}_t (i) \tag{6}$$

where $E_t (i) [\cdot] \equiv E [\cdot | \mathcal{I}_t (i)]$ is agent i's (first-order) expectation conditioned on information available to her in period t (defined below); and $\boldsymbol{v}_t (i)$ is a transitory, mean zero shock specific to agent i in period t (defined below).

Non-zero elements in $\boldsymbol{\lambda}_1$ against higher-order average expectations capture strategic considerations in agents' actions. Note that the terms in $\boldsymbol{x}_t$ and $\boldsymbol{v}_t (i)$ are included here to make the model as general as possible. They allow for the possibility that components of i's signal vector may have direct economic significance in addition to their informational role. Note, too, that although $\boldsymbol{x}_t$ may be included in agents' decision rule, it is *not* directly observed.[21]

Equation (6) nests a wide array of commonly studied settings and its derivation will invariably be context-specific. For example, in the model of Morris and Shin (2002), (6) would be written as

$$g_t (i) = (1-\beta) \left[1 \quad \beta \quad \beta^2 \quad \cdots\right] E_t (i) \left[\overline{x}_t^{(0:\infty)}\right] \tag{7}$$

where β is the weight placed on average actions. This example is explored further in section 4 below. Another example could be the setting of firms' prices, with strategic interaction arising from firms' demand schedules being a function of their relative prices.

Agents' information

Agents possess common knowledge of joint rationality, in the sense of Nimark (2008), so that they are aware of the structure and the coefficients of the system. Their information sets then evolve as:

$$\mathcal{I}_0 (i) = \{\Omega, \Phi\} \qquad \mathcal{I}_t (i) = \{\mathcal{I}_{t-1} (i), \boldsymbol{s}_t (i)\} \tag{8}$$

where Ω is the set of all system coefficients, $\boldsymbol{s}_t (i)$ is the signal vector received each period and $\Phi : [0,1] \rightarrow [0,1]$ is the (cumulative) distribution from which agents' observation targets in the

[21] For example, a firm may privately observe their productivity, which includes both aggregate and idiosyncratic components, but not their separate values. If their decision rule relies directly on their productivity, it will include a term in the aggregate productivity even though firms do not observe it directly.

network are drawn, assumed to be identical and independent, both across agents and across time. $\Phi(i)$ is absolutely continuous over the range $[0,1]$ and has p.d.f. $\phi(i)$.

Each agent's signal vector is made up of two, distinct components. First, a combination of *public and private signals* based on the current underlying state or the lagged full state. These are essentially identical to those used in existing incomplete information work such as Nimark (2008). Second, each agent receives a *social signal* vector derived from observing competitors' actions over the network with a one-period lag:

$$\begin{aligned} \boldsymbol{s}_t(i) &= \begin{bmatrix} \boldsymbol{s}_t^p(i) \\ \boldsymbol{s}_t^s(i) \end{bmatrix} \\ \boldsymbol{s}_t^p(i) &= D_1 \boldsymbol{x}_t + D_2 X_{t-1} + R_1 \boldsymbol{v}_t(i) + R_2 \boldsymbol{e}_t + R_3 \boldsymbol{z}_{t-1} \\ \boldsymbol{s}_t^s(i) &= \boldsymbol{g}_{t-1}(\boldsymbol{\delta}_{t-1}(i)) \end{aligned} \tag{9}$$

Public and private signals may include both current and lagged information and are noisy, including three sources of uncertainty:

- $\boldsymbol{v}_t(i)$ is a vector of transitory shocks specific to agent i in period t, drawn from independent and identical Gaussian distributions with mean zero and variance Σ_{vv}. These may simply be noise in agents' private signals or may carry economic significance, depending on the context.
- $\boldsymbol{z}_t$ is a vector of *network shocks* (see equation 17 below), comprised of weighted sums of all agents' idiosyncratic shocks.
- $\boldsymbol{e}_t$ is a vector of transitory "noise" shocks to public signals, drawn from an independent Gaussian distribution with mean zero and variance Σ_{ee}.

Although agents may observe signals based on the current underlying state ($\boldsymbol{x}_t$), they do not observe signals based on the current hierarchy of expectations about the state (X_t). This is because to do so would involve agents observing a signal based on their beliefs before they have formed them!

Terms in X_{t-1} and $\boldsymbol{z}_{t-1}$ are permitted (instead of just $\boldsymbol{x}_{t-1}$) to allow agents to observe aggregate variables with a lag,[22] and thus the past effect of their network learning.

Social signals are observations of the previous-period actions of specific agents, with the function $\boldsymbol{\delta}_t$ mapping each agent onto their observation targets:

$$\boldsymbol{\delta}_t : [0,1] \to [0,1]^q \tag{10}$$

where q is the number of agents observed. In other words, $\boldsymbol{\delta}_t(i)$ is the result of i's q separate draws from Φ for period t. For presentational simplicity, I will typically assume that $q = 1$ (i.e. that all agents observe a single other agent) and simply write $j = \delta_t(i)$ to mean that agent j's period-t action will be observed by agent i (in period $t+1$) so that on the i^{th} row of the network adjacency matrix, we will have $G_{i,\delta_t(i)} = 1$.

To speak of the observee of an observee, one may write $\delta_s(\delta_t(i))$: the identity of the agent whose period-s action is observed by the agent whose period-t action is observed by agent i.

[22] For example, if allocations are functions of the entire hierarchy of beliefs, then the publication by a national statistical organisation of an estimate of the previous period's GDP would be a function of both X_{t-1} and $\boldsymbol{z}_{t-1}$.

With agent i observing the previous-period action of a single competitor, their social signal is therefore given by

$$\begin{aligned} \boldsymbol{s}_t^s(i) &= g_{t-1}(\delta_{t-1}(i)) \\ &= \boldsymbol{\lambda}_1' E_{t-1}(\delta_{t-1}(i))[X_{t-1}] + \boldsymbol{\lambda}_2' \boldsymbol{x}_{t-1} + \boldsymbol{\lambda}_3' \boldsymbol{v}_{t-1}(\delta_{t-1}(i)) \end{aligned} \tag{11}$$

3.2 The observation network

Because agent i's social signal is based on her observee's expectation, Bayesian updating then requires that i include $E_t(\delta_t(i))[X_t]$ in her own state vector of interest. However, knowing that agent $\delta_t(i)$ is himself considering $\delta_t(\delta_t(i))$ then requires that i also maintain an estimate of $E_t(\delta_t(\delta_t(i)))[X_t]$, and so forth. This is the explosion of the state vector in p (the number of compound expectations) described in section 2.2 above. In order to make the problem tractable, I make two key assumptions:

Assumption 1. *The network is stochastic and opaque, in that:*

- *all agents observe the same number of other agents;*
- *observees are drawn from identical, fully independent distributions with p.d.f. $\phi(i)$;*
- *agents know the identities of the other agents they observe;*
- *agents do not know who they are observed by; and*
- *agents do not learn about the network topology over time.*

To obtain this last point, I suppose that agents make a fresh draw of whom to observe every period, in which case nothing *could* be learned about the network topology (since it changes every period).

Assumption 2. *The network is asymptotically irregular, in that its degree sequence is asymptotically non-uniform (see definition 3).*

As shown in section 2.3 and expanded on below, assumption 2 is sufficient to ensure that idiosyncratic shocks do not "wash out" in aggregation, but will instead enter into agents' average beliefs.

Note, too, that the unconditional expected (in) degree of agent i in a network of n agents will be $E_n[d_i] = q\phi_n(i)$, so that $E_n[d_i] \to 0$ as $n \to \infty$.

3.3 Agents' signal extraction problem

It will be shown below that the hierarchy of agents' expectations obeys the following vector ARMA(1,1) law of motion:

$$X_t \equiv \mathbb{E}_t^{(0:\infty)}[\boldsymbol{x}_t] = FX_{t-1} + G_1\boldsymbol{u}_t + G_2\boldsymbol{z}_t + G_3\boldsymbol{e}_t + G_4\boldsymbol{z}_{t-1} \tag{12}$$

where $\boldsymbol{z}_t$ is a vector of transitory *network shocks*, derived as weighted sums of agents' idiosyncratic shocks. The exact statistical properties of $\boldsymbol{z}_t$ are derived below in proposition 1.

The system described here is not in the form of a classic state space problem, however, both because of the presence of the lagged state in agents' signals (9) and because of the moving average component of the law of motion (12). The most common approach to addressing these features is to stack the state vector with both its own lag and the lag of the shock with the moving average component, thus creating a combined state that follows an VAR(1) process:

$$\begin{bmatrix} X_t \\ \boldsymbol{z}_t \\ X_{t-1} \end{bmatrix} = \begin{bmatrix} F & G_4 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ I & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} X_{t-1} \\ \boldsymbol{z}_{t-1} \\ X_{t-2} \end{bmatrix} + \begin{bmatrix} G_1 & G_2 & G_3 \\ \mathbf{0} & I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_t \\ \boldsymbol{z}_t \\ \boldsymbol{e}_t \end{bmatrix}$$

and then to express agents' signals in terms of this combined state and estimate the system as a classic filtering problem. This approach more than doubles the size of the state vector, though, which may present problems when simulating the system with finite computing resources (and particularly so in the present setting with multiple compound expectations). Fortunately, the following lemma grants us that it is not necessary here to include $\boldsymbol{z}_t$ in the state vector of interest.[23]

Lemma 1. *Agents' contemporaneous expectations of the network shocks are zero:*

$$E_t(i)\left[\boldsymbol{z}_t\right] = \mathbf{0}\ \forall i, t \tag{13}$$

Since all individual agents' expectations of the network shock are zero, it must be the case that all average expectations (simple or weighted) of the network shock are also zero and since agents are jointly rational, this must be common knowledge. There is therefore no need to include any expectation of $\boldsymbol{z}_t$ within the state vector to be estimated.

Because of the linearity of the system, the best linear estimator in the sense of minimising the mean squared error[24] will be a Kalman filter:[25]

$$E_t(i)\left[X_t\right] = E_{t-1}(i)\left[X_t\right] + K\left\{\boldsymbol{s}_t(i) - E_{t-1}(i)\left[\boldsymbol{s}_t(i)\right]\right\} \tag{14}$$

where K is a time-invariant projection matrix (the Kalman gain). As in other models of imperfect common knowledge, since X_t includes $\overline{\boldsymbol{x}}_{t|t}^{(0:\infty)}$, it must be that (a) the state vector to be estimated is of infinite dimension; and (b) the Kalman filter serves a dual role, both as estimator and as part of the law of motion for the state vector.

In the context of firms' price-setting behaviour, Nimark (2008) allowed agents to observe an aggregate signal (the average price) from the previous period in addition to their private signals. This means that each agent's signal vector includes a linear combination of the *entire hierarchy* of previous-period expectations. As a result, the solution must be found for all higher-order expectations simultaneously and the state vector of interest expands to include $\overline{\boldsymbol{x}}_{t-1|t-1}^{(0:\infty)}$ so that $X_t = \left[\overline{\boldsymbol{x}}_{t|t}^{(0:\infty)\prime} \quad \overline{\boldsymbol{x}}_{t-1|t-1}^{(0:\infty)\prime}\right]'$.

An alternative to including $\overline{\boldsymbol{x}}_{t-1|t-1}^{(0:\infty)}$ in the state vector of interest is to retain the current signal vector and instead to modify the Kalman filter:

$$E_t(i)\left[\overline{\boldsymbol{x}}_{t|t}^{(0:\infty)}\right] = K\boldsymbol{s}_t(i) + \left(F - K\left(D_1 F + D_2\right)\right) E_{t-1}(i)\left[\overline{\boldsymbol{x}}_{t-1|t-1}^{(0:\infty)}\right]$$

This approach was first developed by Nimark (2008, 2011b) and is also used in the current paper to avoid the need to stack the state vectors of interest.

[23]Unless otherwise stated, the proof of this and all further propositions may be found in the appendix.

[24]With all shocks drawn from Gaussian distributions, it will be the best such estimator, linear or otherwise.

[25]A derivation of the standard Kalman filter may be found in most texts on dynamic macroeconomics (e.g. Ljungqvist and Sargent (2004)) or time series analysis (e.g. Hamilton (1994)).

3.4 Peering into the mist

In implementing the Kalman filter (14), each agent must form a prior expectation of the signal(s) they will receive in the next period. Stepping equation (11) forward by one period, it is easily seen that it is necessary for agent i to construct $E_t(i)[E_t(\delta_t(i))[X_t]]$ as part of her prior for period $t+1$:

$$\begin{aligned} E_t(i)\left[\boldsymbol{s}^s_{t+1}(i)\right] &= E_t(i)\left[g_t(\delta_t(i))\right] \\ &= E_t(i)\left[\boldsymbol{\lambda}_1' E_t(\delta_t(i))[X_t] + \boldsymbol{\lambda}_2'\boldsymbol{x}_t + \boldsymbol{\lambda}_3'\boldsymbol{v}_t(\delta_t(i))\right] \end{aligned}$$

Constructing $E_t(i)[E_t(\delta_t(i))[X_t]]$ requires, in turn, that agent i take a view regarding who $\delta_t(i)$ is observing: that is, the action of $\delta_{t-1}(\delta_t(i))$.

Lemma 2. *Given assumption 1 and common knowledge of rationality, agents' use of a linear estimator implies that all agents treat all other agents* as though *they observe a common, weighted average of previous-period actions, with the weights given by the distribution ϕ.*

From equation (6), it follows that the weighted-average action, $\widetilde{g}_t$, is given by:

$$\begin{aligned} \widetilde{g}_t &\equiv \int_0^1 g_t(j)\,\phi(j)\,dj \\ &= \boldsymbol{\lambda}_1'\widetilde{E}_t[X_t] + \boldsymbol{\lambda}_2'\boldsymbol{x}_t + \boldsymbol{\lambda}_3'\widetilde{\boldsymbol{v}}_t \end{aligned} \tag{15}$$

It is not possible, in general, to make use of some law of large numbers to disregard the effect of idiosyncratic shocks in the weighted-average action – that is, one cannot assume that $\widetilde{\boldsymbol{v}}_t \equiv \int_0^1 \boldsymbol{v}_t(j)\,\phi(j)\,dj$ will be equal to zero – because the weights applied to each agent may not be sufficiently close to equal. As an extreme example, if all agents were to observe agent 1 and nobody else (i.e. $\phi(1) = 1$ and $\phi(i) = 0\ \forall i \neq 1$), then $\widetilde{\boldsymbol{v}}_t = \boldsymbol{v}_t(1)$, which in any given period will be non-zero, almost surely.

Equation (15) is used for consideration of agents that are one step away in the observation network, but it is also necessary to consider the actions of agents that are two or more steps away:

Definition 4. *Let $\delta_t(i)$ be a period-t mapping from agent i to their observation target, drawn from the distribution Φ_n. The* ***p*** $^{\boldsymbol{th}}$***-weighted average*** *of agents' idiosyncratic shocks is given by*

$$\overset{p:\sim}{\boldsymbol{v}}_{n,t} \equiv \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t}_{p}(i)))\right) \qquad \overset{p:\cdot\cdot}{\boldsymbol{v}}_{n,t} \equiv \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t}_{p-1}(i)))\right)\phi_n(i) \tag{16}$$

For example, the 1^{st}-weighted average is $\overset{1:\sim}{\boldsymbol{v}}_{n,t} \equiv \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{v}_t(\delta_t(i))$ and represents the average idiosyncratic shock over all agents' observees. The 2^{nd}-weighted average is $\overset{2:\sim}{\boldsymbol{v}}_{n,t} \equiv \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{v}_t(\delta_t(\delta_t(i)))$ and represents the average idiosyncratic shock over all agents' observees' observees.

The following proposition then demonstrates an equivalence between $\overset{p:\sim}{\boldsymbol{v}}_{n,t}$ and $\overset{p:\cdot\cdot}{\boldsymbol{v}}_{n,t}$ as the number of agents approaches infinity and characterises the resultant distributions:

Proposition 1. *Suppose that $\boldsymbol{v}_t(i) \sim$ i.i.d. $N(\boldsymbol{0}, \Sigma_{vv})\ \forall i, t$ and assumptions 1 and 2 hold. Then in the limit (as $n \to \infty$):*

1. $\overset{p:\sim}{\boldsymbol{v}}_{n,t} \xrightarrow{d} \overset{p:\sim}{\boldsymbol{v}}_t \quad \forall p \geq 1$ *where* $\overset{p:\sim}{\boldsymbol{v}}_t \sim N\left(\boldsymbol{0}, \Sigma^{\{p\}}_{\widetilde{v}\widetilde{v}}\right)$ *and* $\Sigma^{\{p\}}_{\widetilde{v}\widetilde{v}} = (1-(1-\zeta^*)^p)\,\Sigma_{vv}$
2. $Cov\left(\overset{p:\sim}{\boldsymbol{v}}_t, \overset{r:\sim}{\boldsymbol{v}}_t\right) = \Sigma^{\{p\}}_{\widetilde{v}\widetilde{v}} \quad \forall p < r$

3. ${}^{p:\cdot\cdot}\boldsymbol{v}_{n,t} \xrightarrow{\mathcal{L}^2} {}^{p:\sim}\boldsymbol{v}_t \quad \forall p \geq 1$

It may also be worth noting that the variance of the p^{th}-weighted average may also be expressed recursively as $\Sigma^{\{p\}}_{\underset{\sim}{vv}} = \zeta^* \Sigma_{vv} + (1 - \zeta^*) \Sigma^{\{p-1\}}_{\underset{\sim}{vv}}$. The following two corollaries then trivially follow:

Corollary 1. *$\Sigma_{vv} \geq \cdots \geq \Sigma^{\{3\}}_{\underset{\sim}{vv}} \geq \Sigma^{\{2\}}_{\underset{\sim}{vv}} \geq \Sigma^{\{1\}}_{\underset{\sim}{vv}}$ where $\geq$ is in the sense that the difference between the two is a positive-definite matrix.*

Corollary 2. *$E\left[{}^{p:\sim}\boldsymbol{v}_t \,|\, {}^{1:\sim}\boldsymbol{v}_t = \boldsymbol{a}\right] = \boldsymbol{a} \; \forall p \geq 2$*

The first of these is a necessary component of approximating the full solution with a finite state vector (see section 3.6 below) and the latter is used when simulating the effects of network learning.

The increasing variance of higher-weighted averages captures the effect of what Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Saleh (2012) refer to as the p^{th}-order *interconnectivity* of the network, with each step back in the network more and more likely to arrive at the most highly connected agents. I avoid their nomenclature, however, and reserve the word "order" to refer to higher-order expectations.

Definition 5. *The vector of* ***network shocks****, $\boldsymbol{z}_t$, is the infinite sequence of higher-weighted averages of agents' idiosyncratic shocks:*

$$\boldsymbol{z}_t \equiv \begin{bmatrix} {}^{1:\sim}\boldsymbol{v}_t \\ {}^{2:\sim}\boldsymbol{v}_t \\ {}^{3:\sim}\boldsymbol{v}_t \\ {}^{4:\sim}\boldsymbol{v}_t \\ \vdots \end{bmatrix} \sim N(\boldsymbol{0}, \Sigma_{zz}) \qquad \Sigma_{zz} = \begin{bmatrix} \Sigma^{\{1\}}_{\underset{\sim}{vv}} & \Sigma^{\{1\}}_{\underset{\sim}{vv}} & \Sigma^{\{1\}}_{\underset{\sim}{vv}} & \Sigma^{\{1\}}_{\underset{\sim}{vv}} & \cdots \\ \Sigma^{\{1\}}_{\underset{\sim}{vv}} & \Sigma^{\{2\}}_{\underset{\sim}{vv}} & \Sigma^{\{2\}}_{\underset{\sim}{vv}} & \Sigma^{\{2\}}_{\underset{\sim}{vv}} & \cdots \\ \Sigma^{\{1\}}_{\underset{\sim}{vv}} & \Sigma^{\{2\}}_{\underset{\sim}{vv}} & \Sigma^{\{3\}}_{\underset{\sim}{vv}} & \Sigma^{\{3\}}_{\underset{\sim}{vv}} & \cdots \\ \Sigma^{\{1\}}_{\underset{\sim}{vv}} & \Sigma^{\{2\}}_{\underset{\sim}{vv}} & \Sigma^{\{3\}}_{\underset{\sim}{vv}} & \Sigma^{\{4\}}_{\underset{\sim}{vv}} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{17}$$

Including these higher weighted averages is necessary because of the recursive nature of agents' learning through the Kalman filter: it will be shown below that ${}^{1:\sim}E_t[X_t]$ is a function of ${}^{1:\sim}\boldsymbol{v}_t$ and ${}^{2:\sim}E_{t-1}[X_{t-1}]$, while ${}^{2:\sim}E_t[X_t]$ is a function of ${}^{2:\sim}\boldsymbol{v}_t$ and ${}^{3:\sim}E_{t-1}[X_{t-1}]$, etc.

3.5 Social learning over an opaque, irregular network

I am now in a position to present the main result of this paper.

Theorem 1. *Given the broad setting described above and assumptions 1 and 2, the hierarchy of agents' aggregate expectations will obey the following VARMA(1,1) law of motion:*

$$X_t \equiv \begin{bmatrix} \boldsymbol{x}_t \\ \overline{E}_t[X_t] \\ {}^{1:\sim}E_t[X_t] \\ {}^{2:\sim}E_t[X_t] \\ \vdots \end{bmatrix} = FX_{t-1} + G_1\boldsymbol{u}_t + G_2\boldsymbol{z}_t + G_3\boldsymbol{e}_t + G_4\boldsymbol{z}_{t-1}$$

$$\text{where} \quad \overline{E}_t[\cdot] = \int_0^1 E_t(i)[\cdot]\,di$$

$${}^{1:\sim}E_t[\cdot] = \int_0^1 E_t(\delta_t(i))[\cdot]\,di$$

$${}^{2:\sim}E_t[\cdot] = \int_0^1 E_t(\delta_t(\delta_t(i)))[\cdot]\,di$$

$$\vdots$$

and with each successively higher-weighted average expectation having a smaller effect on the simple-average expectation.

Although the complete derivation is provided in the appendix, an outline of the agents' learning process may be of interest. To begin, I define the general notation that $\theta^{\text{err}}_{t|q}(i)$ represents the *error* in agent i's period-q expectation regarding θ_t. In particular, the following will be used:

$$\begin{aligned} \boldsymbol{s}^{\text{err}}_{t|t-1}(i) &\equiv \boldsymbol{s}_t(i) - E_{t-1}(i)[\boldsymbol{s}_t(i)] && : \text{signal innovation} \\ X^{\text{err}}_{t|t-1}(i) &\equiv X_t - E_{t-1}(i)[X_t] && : \text{prior expectation error} \\ X^{\text{err}}_{t|t}(i) &\equiv X_t - E_t(i)[X_t] && : \text{contemporaneous expectation error} \end{aligned}$$

The filter

As with a standard Kalman filter, the Kalman gain is calculated as:

$$K_t = Cov(X_t, \boldsymbol{s}^{\text{err}}_{t|t-1}(i))\left[Var\left(\boldsymbol{s}^{\text{err}}_{t|t-1}(i)\right)\right]^{-1} \tag{18}$$

where $\boldsymbol{s}^{\text{err}}_{t|t-1}(i)$ is the agent's signal innovation (the portion of their signal that was not forecastable). Under a classic filtering problem with no network learning, the agents' signal innovation is a function of their expectation error from the previous period and current period shocks:

$$\boldsymbol{s}^{\text{err}}_{t|t-1}(i) = M_1 X^{\text{err}}_{t-1|t-1}(i) + N_1\boldsymbol{u}_t + N_2\boldsymbol{v}_t(i) + N_3\boldsymbol{e}_t$$

In contrast, with network learning the signal innovation is expressed as follows (the additional terms are shown in red):

$$\begin{aligned} \boldsymbol{s}^{\text{err}}_{t|t-1}(i) = {} & M_1 X^{\text{err}}_{t-1|t-1}(i) + M_2 X^{\text{err}}_{t-1|t-1}(\delta_{t-1}(i)) + M_3 X_{t-1} \\ & + N_1\boldsymbol{u}_t + N_2\boldsymbol{v}_t(i) + N_3\boldsymbol{e}_t + N_4\boldsymbol{v}_{t-1}(\delta_{t-1}(i)) + N_5\boldsymbol{z}_{t-1} \end{aligned} \tag{19}$$

Note that innovation in i's signal includes not only a term in their own previous-period expectation error but also a term in their *observee*'s expectation error. As such, both the covariance and variance terms in the Kalman gain (18) will therefore include terms in both the variance of i's expectation error, $V_{t-1|t-1} \equiv E\left[X^{\text{err}}_{t-1|t-1}(i)\,X^{\text{err}}_{t-1|t-1}(i)'\right]$, and the *covariance* between any two agents' errors, $W_{t-1|t-1} \equiv E\left[X^{\text{err}}_{t-1|t-1}(i)\,X^{\text{err}}_{t-1|t-1}(j)'\right]$.

The variance in agents' own expectation errors then updates in via the familiar Riccati equation, but since the variance-covariance of $\boldsymbol{s}^{\text{err}}_{t|t-1}(i)$ includes terms in $W_{t-1|t-1}$, a corresponding expression must also be found for updating the covariance between agents' errors.

The law of motion

The law of motion starts from the basic form of the Kalman filter:

$$E_t(i)[X_t] = FE_{t-1}(i)[X_{t-1}] + K_t\boldsymbol{s}^{\text{err}}_{t|t-1}(i)$$

Equation (19) is substituted in for $\boldsymbol{s}^{\text{err}}_{t|t-1}(i)$ and a simple average is taken to obtain $\overline{E}_t[X_t]$. Since the signal innovation includes a term in $E_{t-1}(\delta_{t-1}(i))[X_{t-1}]$ (from the observee's expectation error), taking the simple average over i turns this into a term in $\overset{1:\sim}{E}_{t-1}[X_{t-1}]$, thereby introducing the need to also determine the (first) weighted-average expectation.

Taking the weighted average of the filter to obtain $\overset{1:\sim}{E}_t[X_t]$ then produces a term in $\overset{2:\sim}{E}_{t-1}[X_{t-1}]$, thus requiring that the 2^{nd}-weighted average expectation be included. The 2^{nd}-weighted average expectation subsequently produces a term in the 3^{rd}-weighted average expectation, and so forth.

3.6 Working with a finite approximation

The full state vector of interest and, hence, the transition matrices in the law of motion and the filter variances in the Kalman filter are all of infinite dimension. The full solution therefore cannot be found in practice and must be approximated with a truncated state.

Proposition 2. *An arbitrarily accurate approximation of the full solution implied by theorem 1 may be obtained by defining cut-off on the number of higher orders of expectation, k^*, and the number of higher-weighted compound expectations, p^*; and including all weights and all orders from zero up to these cut-offs.*

Note that the size of the state vector can still be very large even when operating with few state variables and quite low choices of k^* and p^*. Table 1 lists the sizes that emerge for a variety of choices.

m	k^*	No network (standard ICK)	With network learning ($p^* = 3$)
1	4	5×5 : 200 B	121×121 : 114.4 KB
1	6	7×7 : 392 B	1093×1093 : 9.1 MB
4	4	16×16 : 2.0 KB	484×484 : 1.8 MB
4	6	28×28 : 6.1 KB	4372×4372 : 145.8 MB

Table 1: Size (each) of F, U, V and W, assuming use of double-precision.

Given the size of the matrices involved, problems of *numerical instability* must be considered when implementing the model. When iterating a large system over many steps, round-off errors that necessarily occur with floating-point operations on computers can accumulate and magnify to the extent that the system does not converge. Such a problem is, regrettably, relatively common in the implementation of larger Kalman filters and typically first appears as a failure of symmetry or positive definiteness in the variance matrices of the Ricatti equation.

Arguably the most robust (to roundoff error) implementations of Kalman filters are those that factor the relevant variance matrices, with a modified Cholesky decomposition (a "UD decomposition") the most commonly used.[26] Using this technique for a regular Kalman filter, the algorithm for implementing the temporal update of the filter (from $V_{t-1|t-1}$ to $V_{t|t-1}$) was developed by Thornton (1976) and that for the observational update (from $V_{t|t-1}$ to $V_{t|t}$) by Bierman (1977).

Unfortunately, although the model developed here is amenable to use of the Thornton temporal update, the Bierman observational update algorithm is not applicable. This is because the inclusion of social signals introduces the need to consider the covariance of agents' expectation errors so that, when calculating the Kalman gain, the covariance between the state (X_t) and the signal innovation ($\boldsymbol{s}^{\text{err}}_{t|t-1}(i)$) can no longer be expressed in the form

$$Cov\left(X_t, \boldsymbol{s}^{\text{err}}_{t|t-1}(i)\right) = V_{t|t-1}H$$

which is required for Bierman's factorisation. A successful UD implementation of the current model would therefore require the derivation of a new algorithm in the style of Bierman that accounted for the more complex structure of the Kalman gain found here. This is left for future research.

[26] A UD decomposition breaks V into UDU' with U unit upper triangular (i.e. with ones on the leading diagonal) and D diagonal. By working exclusively on these component matricies, the implied variance matricies remain well defined.

4 An illustrative example

I here present a simplified example to illustrate some of the results that emerge from adding network learning to a setting of strategic complementarity. Key simplifying assumptions include:

- A univariate underlying state
- No public signals
- No lagged signals (except the social signal through the network)
- Private signals serve an information role only
- Agents are myopic, in that they optimise on a period-by-period basis

Section 4.4 below extends this model to inclusion of a lagged public signal, while section 4.5 illustrates how to apply the model to dynamic settings where agents are forward looking in their decision rule.

4.1 The simplified model

There exists only a single hidden state that follows an AR(1) process and about which agents each observe a single, unbiased private signal

$$\begin{aligned} x_t &= \rho x_{t-1} + u_t & u_t &\sim N\left(0, \sigma_u^2\right) \\ s_t^p(i) &= x_t + v_t(i) & v_t(i) &\sim N\left(0, \sigma_v^2\right) \end{aligned}$$

with u_t and $v_t(i)$ being fully independent for all i and t. Agents face quadratic losses from mismatch between their action, a single hidden state and the average action of others:[27]

$$u_i(\boldsymbol{g}_t, x_t) = -(1-\beta)\left[(g_t(i) - x_t)^2\right] - \beta\left[(g_t(i) - \overline{g}_t)^2\right] \quad \beta \in (0,1)$$

With agents maximising their expected payoff without explicitly knowing the state or the average action that other agents will take, their optimal action is given by

$$g_t(i) = (1-\beta) E_t(i)[x_t] + \beta E_t(i)[\overline{g}_t]$$

Taking the simple average of this and repeatedly substituting it back in, we obtain

$$g_t(i) = (1-\beta)\begin{bmatrix} 1 & \beta & \beta^2 & \cdots \end{bmatrix} E_t(i)\left[\overline{x}_{t|t}^{(0:\infty)}\right]$$

With each agent observing the previous-period action of q competitors, theorem 1 then grants that the following laws of motion emerge:

$$\begin{aligned} \boldsymbol{x}_t &= \rho\, \boldsymbol{x}_{t-1} & &+ \boldsymbol{u}_t \\ \overline{E}_t[X_t] &= B\, \boldsymbol{x}_{t-1} + C\, \overline{E}_{t-1}[X_{t-1}] + D \overset{1:\sim}{E}_{t-1}[X_{t-1}] &&+ H\boldsymbol{u}_t \\ \overset{1:\sim}{E}_t[X_t] &= B\, \boldsymbol{x}_{t-1} + C \overset{1:\sim}{E}_{t-1}[X_{t-1}] + D \overset{2:\sim}{E}_{t-1}[X_{t-1}] &&+ H\boldsymbol{u}_t + Q \overset{1:\sim}{\boldsymbol{v}}_t \\ \overset{2:\sim}{E}_t[X_t] &= B\, \boldsymbol{x}_{t-1} + C \overset{2:\sim}{E}_{t-1}[X_{t-1}] + D \overset{3:\sim}{E}_{t-1}[X_{t-1}] &&+ H\boldsymbol{u}_t + Q \overset{2:\sim}{\boldsymbol{v}}_t \\ &\vdots \end{aligned}$$

[27] This utility function is quite common in the network literature (see, e.g., Calvó-Armengol and de Martí, 2007). An alternative utility function described by Morris and Shin (2002) presents the strategic complementarity as being a zero-sum game, but produces the same optimal decision rule for individual agents (although not for a social planner).

$$\text{where}\quad B = \boldsymbol{k}_p\,\rho \qquad\qquad H = \boldsymbol{k}_p$$
$$C = F - BS_x - DT_{w_1} \qquad Q = q\,\boldsymbol{k}_p$$
$$D = q\,\boldsymbol{k}_s\,\boldsymbol{\lambda}_1'$$

with $\boldsymbol{k}_p$ being the Kalman gain applied to the private signal and $\boldsymbol{k}_s$ the Kalman gain applied to each social signal, while S_x and T_{w_1} select x_t and $\overset{1:\sim}{E}_t[X_t]$ respectively from X_t. The transition matrix for the full state therefore takes the following form:

$$F = \begin{bmatrix} \rho & 0 & 0 & 0 & \cdots \\ B & C & D & 0 & \\ B & 0 & C & D & \\ B & 0 & 0 & C & \ddots \\ \vdots & & & & \ddots \end{bmatrix}$$

The p^{th}-weighted expectation is given by:

$$\overset{p:\sim}{E}_t[X_t] = \left(\boldsymbol{k}_p\,\rho S_x + (F - \boldsymbol{k}_p\,\rho S_x)\,T_{w_p}\right)X_{t-1} + q\,\boldsymbol{k}_s\,\boldsymbol{\lambda}_1'\left(T_{w_{p+1}} - T_{w_1}T_{w_p}\right)X_{t-1} + shocks$$

When considering the expectations of agents p levels deep in the network, the component derived from consideration of agents $p+1$ levels deep is captured in the term $q\,\boldsymbol{k}_s\,\boldsymbol{\lambda}_1'\left(T_{w_{p+1}} - T_{w_1}T_{w_p}\right)X_{t-1}$.

In order to simulate the model, it is necessary to form a finite approximation of the solution. For example, if we impose that $k^* = 2$ and $p^* = 2$, then the state vector will have eight elements:

$$X_t = \mathbb{E}_t^{(0:2)}[x_t] = \left[x_t \quad \overline{E}_t[x_t] \quad \overline{E}_t\left[\overline{E}_t[x_t]\right] \quad \overline{E}_t\left[\overset{1:\sim}{E}_t[x_t]\right] \quad \overset{1:\sim}{E}_t[x_t] \quad \overset{1:\sim}{E}_t\left[\overline{E}_t[x_t]\right] \quad \overset{1:\sim}{E}_t\left[\overset{1:\sim}{E}_t[x_t]\right]\right]'$$

It can be readily shown that $Cov\left(\overline{E}_t[X_t], \boldsymbol{s}^{\text{err}}_{t|t-1}(i)\right) = Cov\left(\overset{p:\sim}{E}_t[X_t], \boldsymbol{s}^{\text{err}}_{t|t-1}(i)\right)\ \forall p$, which implies the following repetitive structures for the Kalman gains:

$$\boldsymbol{k}_p = \begin{bmatrix}\kappa_1^p\\ \kappa_2^p\\ \kappa_3^p\\ \kappa_3^p\\ \kappa_2^p\\ \kappa_3^p\\ \kappa_3^p\end{bmatrix} \qquad \boldsymbol{k}_s = \begin{bmatrix}\kappa_1^s\\ \kappa_2^s\\ \kappa_3^s\\ \kappa_3^s\\ \kappa_2^s\\ \kappa_3^s\\ \kappa_3^s\end{bmatrix}$$

and the following coefficients in the law of motion:

$$B = \begin{bmatrix}\kappa_1^p\rho\\ \kappa_2^p\rho\\ \kappa_2^p\rho\end{bmatrix} \quad C = \begin{bmatrix}(1-\kappa_1^p)\,\rho & 0 & -q\kappa_1^s\,(1-\beta)\\ (\kappa_2^p - \kappa_1^p)\,\rho & (1-\kappa_1^p)\,\rho & q\,(\kappa_1^s - \kappa_2^s)\,(1-\beta)\\ (\kappa_2^p - \kappa_1^p)\,\rho & 0 & (1-\kappa_1^p)\,\rho + q\,(\kappa_1^s - \kappa_2^s)\,(1-\beta)\end{bmatrix} \quad D = \begin{bmatrix}q\kappa_1^s\,(1-\beta) & 0 & 0\\ q\kappa_2^s\,(1-\beta) & 0 & 0\\ q\kappa_2^s\,(1-\beta) & 0 & 0\end{bmatrix}$$

For the simulations that follow, I suppose the following parameters:

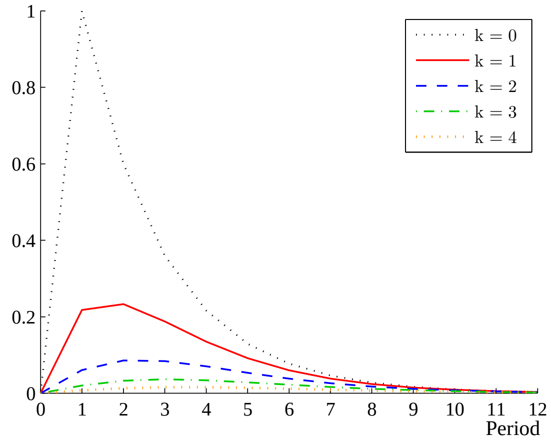
4.2 Aggregate beliefs following a shock to the underlying state

Figure 5 plots impulse responses for the hierarchy of simple-average expectations[28] following a one standard deviation shock to the hidden state, both with and without network learning.

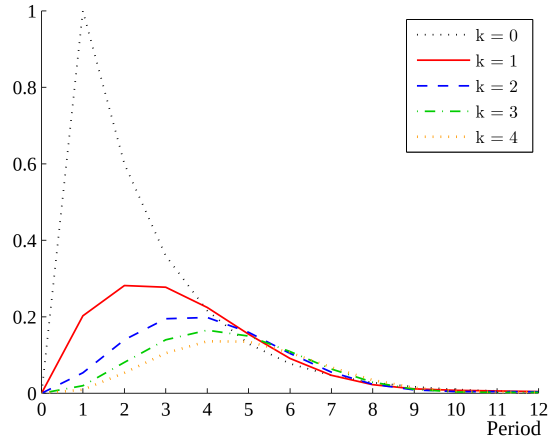
[28] So $k = 0$ denotes the time path of x_t, $k = 1$ the time path of $\overline{E}_t[x_t]$, $k = 2$ the time path of $\overline{E}_t\left[\overline{E}_t[x_t]\right]$ and so on.

Parameter	Value	Description
β	0.5	The relative importance of strategic complementarity
ρ	0.6	The persistence of shocks to the hidden state
σ_v^2/σ_u^2	5.0	The relative innovation variance
ζ^*	0.1	The degree of irregularity in the network

Table 2: Baseline parameterisation



(a) Without network learning ($q = 0$)



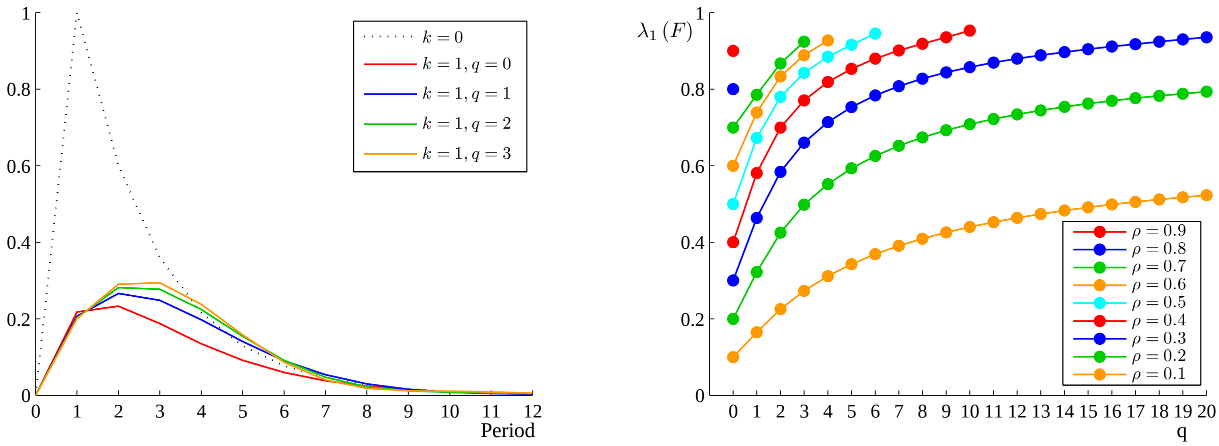
(b) With network learning ($q = 2$)

Figure 5: The hierarchy of simple-average expectations ($\overline{x}_{t|t}^{(0:\infty)}$) following a one standard deviation shock to the underlying state.

Figure 5a shows a standard scenario in the dispersed information literature, with agents only having access to their private signals. Although all agents' signals are unbiased, the presence of noise ensures that they attribute some of their signal to idiosyncratic factors, so the average expectation responds by less than the truth. Since each agent knows this (common knowledge of rationality), each successive order of expectation responds by less than its predecessor. All orders of expectation remain below the the underlying state, so the average expectation error ($\overline{x}_{t|t}^{\text{err}} \equiv x_t - \overline{E}_t[x_t]$) remains strictly positive. The hierarchy of beliefs subsequently decays back to zero with the underlying shock.

Figure 5b then plots the equivalent impulse responses when, in addition to observing their private signals, each agent observes the previous-period action of two competitors. On impact, there is very little difference because social signals are received with a lag (the observation of competitors' actions having been zero in the pre-impact period lowers the beliefs fractionally). In the near term, agents' average expectations are improved relative to the no-network case, with observations of their peers' actions reinforcing their own private signals that an aggregate shock has occurred. In the longer term, however, as the underlying state decays back to zero, agents' beliefs tend to *overshoot* the truth, so the average expectation error ($\overline{x}_{t|t}^{\text{err}}$) becomes negative.

This is herding in the broad sense of Banerjee (1992), but with an amplification from Morris and Shin (2002)-style strategic complementarity. First and most simply, by observing that their competitors' actions were high yesterday, agents infer that the state may be high today. As a result, they partially attribute their low private signals to idiosyncratic noise, consequently choosing a high action themselves. However, although there is no public signal available, by effectively assuming that their competitors all observe the same weighted average action, agents' social observations act as private signals about a public signal that they themselves cannot observe but which they assume is seen by everybody else.[29] For any given agent, their social observation therefore acts as a coordination device for addressing their strategic complementarity concerns. When the underlying state is falling, this therefore acts as a kind of upward bias in social signals for signal extraction purposes.



(a) Simple-average expectations

(b) Largest absolute eigenvalues of F

Figure 6: Varying the number of other agents observed (q)

Figure 6 illustrates the increase in persistence obtained when agents increase the number of competitors they each observe. Figure 6a first plots the truth ($k = 0$) and the first-order simple-average expectations ($k = 1$) for a variety of the number of competitors observed. Figure 6b presents a broader picture, showing the largest absolute eigenvalue of F when agents each observe q competitors for different degrees of persistence in the underlying state (ρ). When there is no network learning ($q = 0$) the full hierarchy of expectations exhibits the same persistence as the underlying state, but as the number of observees increases, the overall persistence of agents' higher-order beliefs rises.

This is **not** to say that the (simple) average expectation becomes arbitrarily persistent. The full hierarchy also includes weighted-average expectations and higher-order expectations of both and it is the higher-weighted, higher-order expectations that exhibit the greatest persistence. Nevertheless, the first-order simple-average expectation, and the hierarchy of simple-average expectations above it, do appear to exhibit greater persistence than the underlying shock in a form of rational herding.

Solutions are not currently able to be found for all combinations of q and ρ, as for more persistent systems (with either high ρ or high q), the filter becomes susceptible to numerical instability.[30] Nevertheless, figure 6b is suggestive of the following conjecture.

Conjecture 1. *Let $\lambda_1(A)$ denote the largest eigenvalue of the transition matrix for the underlying state and let $\lambda_1(F, q)$ be the largest absolute eigenvalue of the transition matrix for the full hierarchy of aggregate expectations when each agent observes q competitors. Then*

- $\lambda_1(F, 0) = \lambda_1(A)$
- $1 > \lambda_1(F, q) > \lambda_1(A) \ \forall q \geq 1$

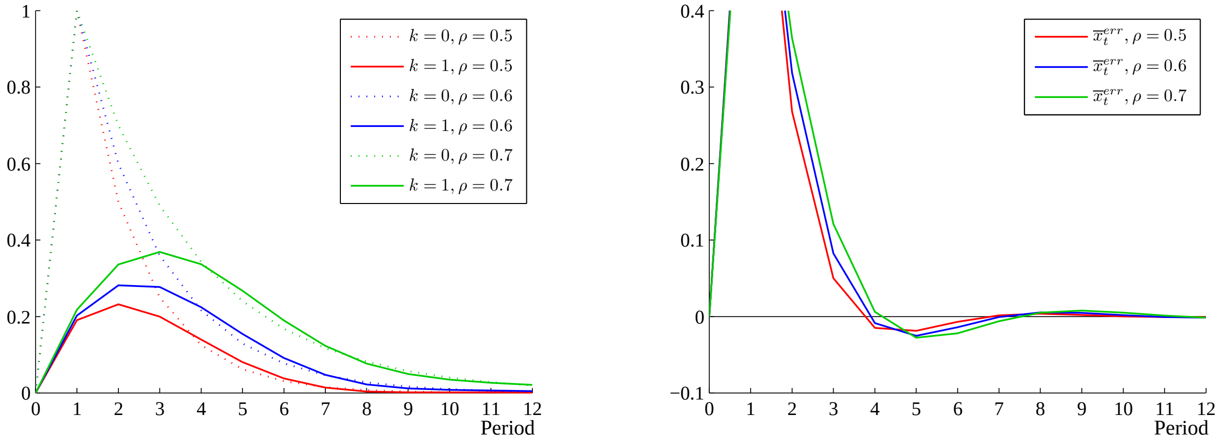
[29] Strictly speaking, agents do not assume that their competitors observe a public signal. Rather, their Bayes-rational signal extraction problem is mathematically equivalent to making the assumption.

[30] Numerical instability is the result of round-off errors due to the limitation of performing floating-point operations on a computer. In Kalman filters, it most often (and, indeed, here) enters predominantly via the Riccati equation updating the variance in expectation errors. The round-off errors cause these variance-covariance matricies to become non-symmetric or non-positive-definite (a mathematical impossibility), which causes the system to explode.

- $\lambda_1(F, q)$ *is strictly increasing in* q *and approaches an asymptote at one*

Proof of this conjecture is left for future work, although it is interesting to note that in contrast to work demonstrating long memory with directed networks (see, for example, Schennach, 2013), the autocorrelation function appears to be absolutely summable here.

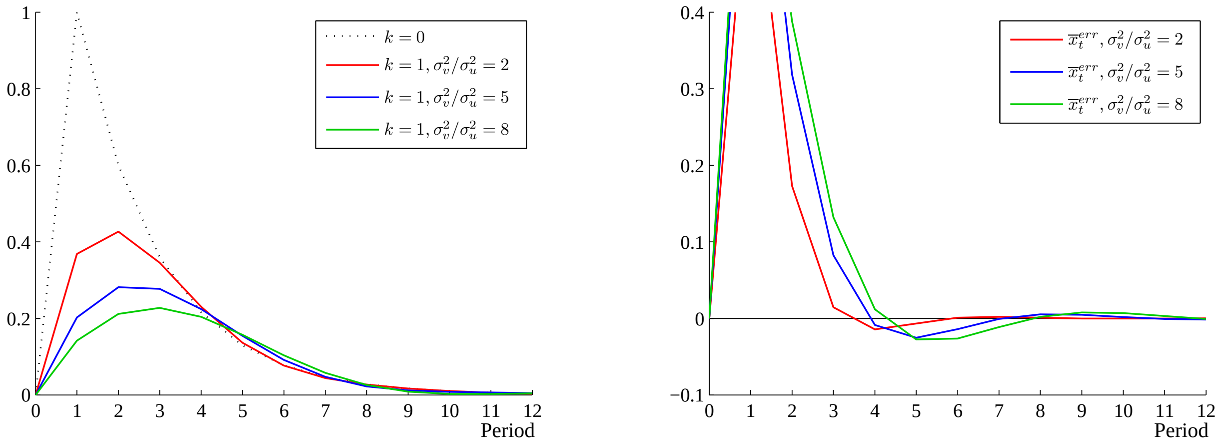
Varying parameters



(a) Simple-average expectations (b) Average expectation errors

Figure 7: Varying underlying persistence (ρ)

Figure 7 shows the impulse responses of first-order simple-average expectations and the corresponding average expectation errors for different values of ρ. Larger values of ρ cause not only larger movements in average expectations, but renders the errors in those expectations larger for longer. In other words, the presence of network learning introduces a persistence multiplier effect so that the persistence of average beliefs increases by more than that of the state.



(a) Simple-average expectations (b) Average expectation errors

Figure 8: Varying the relative innovation variance (σ_v^2/σ_u^2)

Figure 8 then presents equivalent plots for a variety of values for σ_v^2/σ_u^2. Lowering the signal-to-noise ratio of agents' private signals[31] worsens the value of those signals, causing them to rely more

[31] That is, raising the relative variance of idiosyncratic shocks.

heavily on the social signals. This reduces agents' average performance shortly after a shock and produces a stronger overshoot.

4.3 Aggregate beliefs following a network shock

In addition to shocks to the underlying state, the irregularity of the observation network gives rise to the possibility of aggregate *network shocks*: a set of idiosyncratic shocks such that prominent agents happen to draw innovations in one direction (say, positive) while obscure agents draw innovations in the opposite direction. With a continuum of agents, the law of large numbers ensures that the simple average innovation is zero, but weighted averages (with weights given by the probability of being observed) will be non-zero.

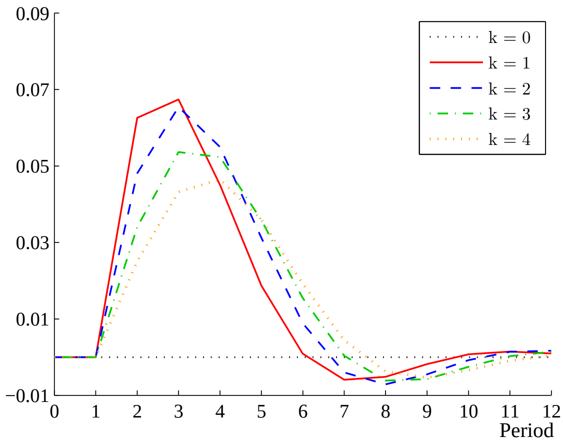
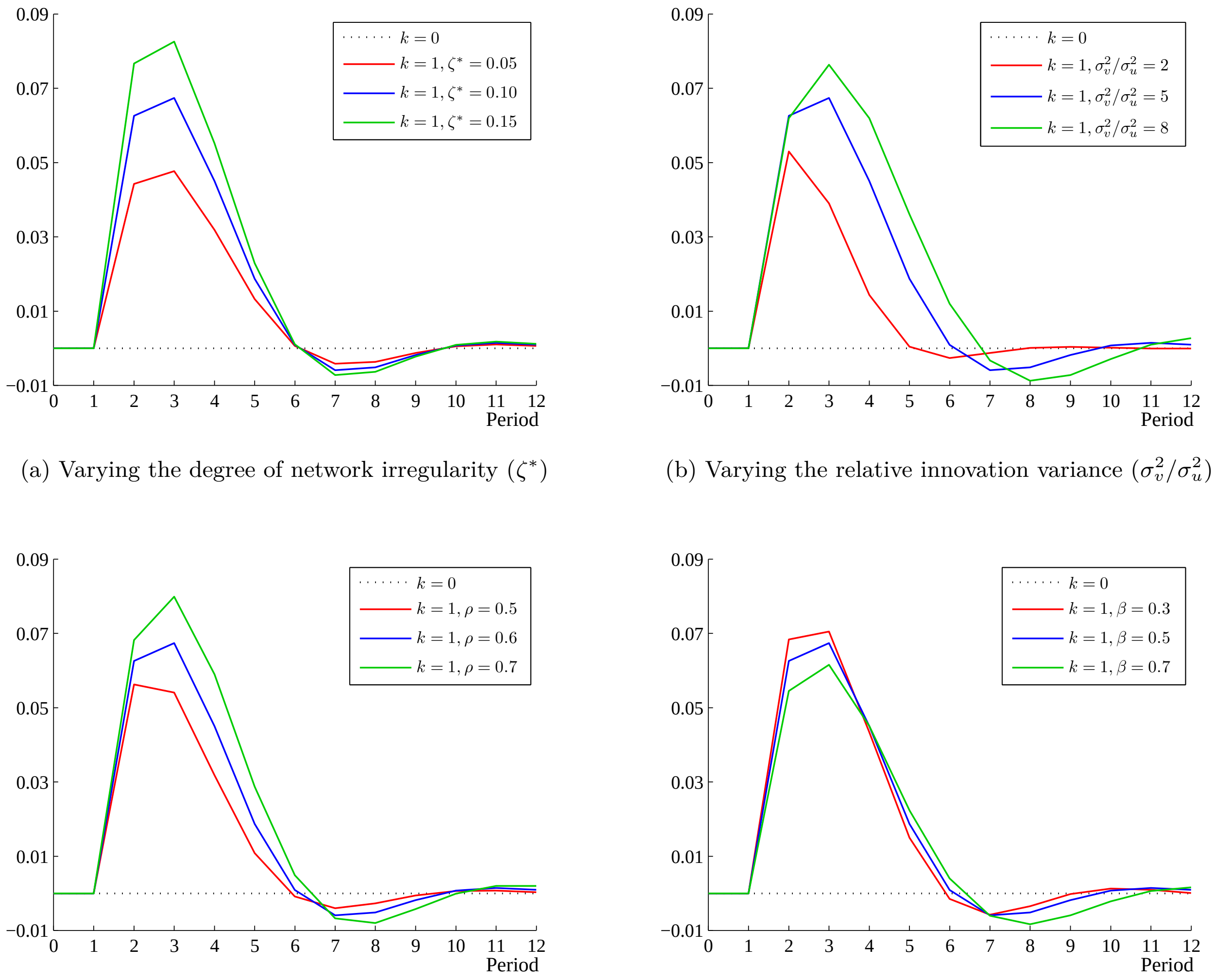


Figure 9: The hierarchy of simple-average expectations ($\overline{x}_{t|t}^{(0:\infty)}$) following a one standard deviation network shock (a one standard deviation shock to ${}^{1:\sim}\boldsymbol{v}_t$ and the corresponding conditional expected value for higher-weighted averages) with agents each observing two competitors ($q = 2$).

Figure 9 plots the hierarchy of simple-average expectations regarding the hidden state following a one standard deviation network shock – strictly, a one standard deviation shock to ${}^{1:\sim}\boldsymbol{v}_t$ plus the corresponding (conditionally) expected value for higher weighted averages – when agents each observe two competitors ($q = 2$). Note that the underlying state remains at zero throughout. Unlike with a shock to the state, there is no movement in aggregate beliefs on impact because the law of large numbers does apply: all agents receive the same social signal from the pre-impact period and movements in the expectations of prominent and obscure agents balance out. In the second period, the average expectation rises as people observe the positive movement in prominent agents' actions from period one and largely ignore the opposite movements by obscure agents. Consequently in period two, despite the average private signal being zero, not just prominent agents but *all* agents, on average, choose positive actions. Aggregate beliefs then gradually decay back to zero as agents continue to receive average private signals of zero but continue to place weight on the previous actions of others.

Overall, the scale of movements in average beliefs is roughly one order of magnitude smaller than those following a true shock to the underlying state. This scale is controlled by the relative variance of the network shocks (recall that $Var\left({}^{1:\sim}\boldsymbol{v}_t\right) = \zeta^* \sigma_v^2$), but also by the persistence of underlying state shocks and the degree of strategic complementarity, as shown in figure 10.

Figure 10a first shows the IRFs of simple-average expectations for different degrees of irregularity in the observation network. At one extreme ($\zeta^* \to 0$), the distribution of links is sufficiently close



(a) Varying the degree of network irregularity (ζ^*)

(b) Varying the relative innovation variance (σ_v^2/σ_u^2)

(c) Varying underlying persistence (ρ)

(d) Varying strategic complementarity (β)

Figure 10: IRFs of simple-average expectations ($\overline{E}_t\,[x_t]$) following a network shock for a variety of parameters ($q = 2$ for all).

to uniform that the law of large numbers applies, meaning that network shocks have no effect. At the other extreme, as the probability of being observed approaches unity for a single agent and zero for everybody else ($\zeta^* \to 1$), that sole agent's idiosyncratic shocks come to play a significant role in shaping average beliefs. Although varying ζ^* changes the magnitude of any movement in agents' average expectations, the profile and persistence of that movement is unchanged.

Figure 10b next shows the effect of network shocks when varying the relative variance of agents' idiosyncratic shocks. As with increasing ζ^*, an increase in σ_v^2/σ_u^2 increases the magnitude of the average expectation's response, but in addition, as seen for shocks to the underlying state above, the increased uncertainty also increases the persistence of the shocks' effects. The profile of the responses also changes, with IRFs become more predominantly hump-shaped for higher relative variances.

Figure 10c then considers the different responses to a network shock for various degrees of persistence in the underlying state. A more persistent hidden state causes a larger and more hump-shaped response to a network shock: (lagged) social signals are more informative about the underlying state when that state is more persistent (higher ρ). Despite the change in profile, the persistence of average beliefs following a network shock does not appear to change with persistence in the underlying state.

Figure 10d finally plots the responses for different degrees of strategic complementarity. Greater strategic complementarity (higher β) dampens the response in period 2 and produces a more pronounced later peak and more persistent response, as agents seek greater confirmation that other agents are adjusting their actions before acting themselves and this flows into the collective signal extraction problem.

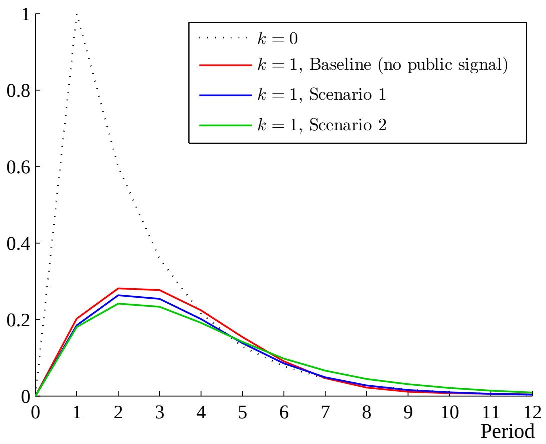
4.4 Adding a (lagged) public signal

It is straightforward to add a public signal to this simplified setting. I consider two cases: one in which the signal is based on the hierarchy of simple-average expectations only and one based on the entire hierarchy (i.e. including weighted-average expectations).

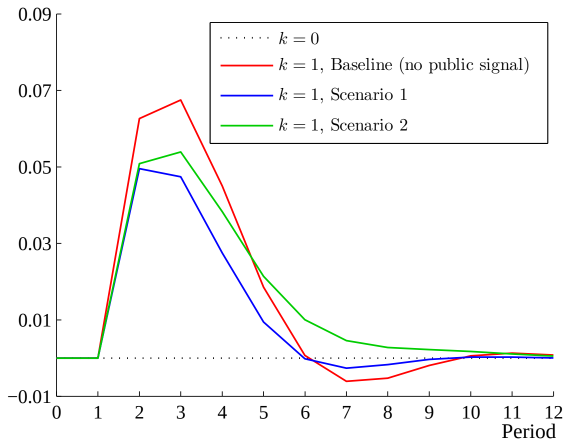
$$\text{Scenario 1: } s_t^{pub} = \mathbf{1}'\overline{x}_{t-1|t-1}^{(0:\infty)} + e_t \tag{20}$$

$$\text{Scenario 2: } s_t^{pub} = \mathbf{1}'X_{t-1} + e_t \tag{21}$$

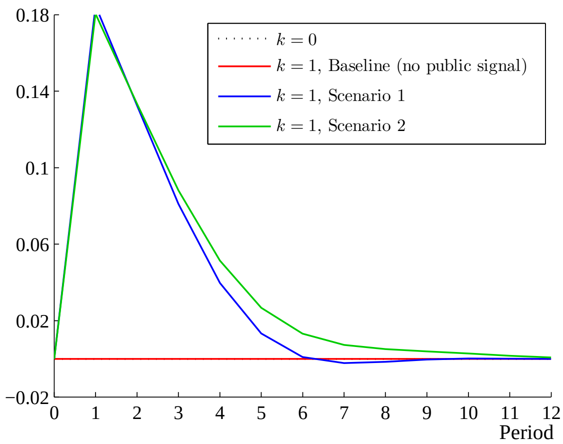
Figure 11 then plots the IRFs of first-order simple-average expectations following a shock to the underlying state, a network shock, or a noise shock in the public signal. A public signal reduces



(a) A shock to the underlying state



(b) A network shock



(c) A public noise shock

Figure 11: IRFs of simple-average expectations ($\overline{E}_t\,[x_t]$) when agents have access to a noisy, lagged public signal ($q = 2$ for all).

the near-term response of simple-average expectations following both shocks to the underlying state and network shocks. For a public signal based only on the hierarchy of simple-average expectations, the overshoot following a shock to the underlying state is reduced (and removed entirely if the signal is perfect). A public signal that is influenced by weighted-average expectations, however, can increase the persistence of simple-average expectations.

4.5 Dynamic actions

The example above presented only a repeated static problem, with agents' optimal actions in period t being only a function of period t variables. This can be extended to a dynamic setting, however. As a simple example, suppose that agents' private signals are given by:

$$\boldsymbol{s}_t^p(i) = B\boldsymbol{x}_t + Q\boldsymbol{v}_t(i)$$

and that the linearised first-order conditions of agents' optimisation problems are given by:

$$g_t(i) = \boldsymbol{\alpha}'\boldsymbol{s}_t^p(i) + \boldsymbol{\eta}_x' E_t(i)[X_t] + \eta_y E_t(i)[\overline{g}_t] + \eta_z E_t(i)[\overline{g}_{t+1}]$$

so that agents' period-t action is a function of their expectation of the average action in both period t and period $t+1$. I show in the appendix that this may be expressed as

$$g_t(i) = \underbrace{(\boldsymbol{\eta}_x' + \eta_y\boldsymbol{a}' + \eta_z\boldsymbol{a}'F)}_{\boldsymbol{\lambda}_1'}E_t(i)[X_t] + \underbrace{\boldsymbol{\alpha}'B}_{\boldsymbol{\lambda}_2'}\boldsymbol{x}_t + \underbrace{\boldsymbol{\alpha}'Q}_{\boldsymbol{\lambda}_3'}\boldsymbol{v}_t(i)$$

where

$$\boldsymbol{a}' \equiv (\boldsymbol{\alpha}'BS + \boldsymbol{\eta}_x'T_s)(I - \eta_y T_s)^{-1}\left(I - \eta_z F T_s (I - \eta_y T_s)^{-1}\right)^{-1}$$

which is clearly in the form of equation (6). Further extensions to consideration of an infinite sum of forward-looking variables in agents' decision rule are straightforward.

5 Conclusion

This paper has introduced and solved a model of social learning over an exogenous directed network with a continuum of agents that satisfies the three requirements that (a) agents are rational; (b) agents act simultaneously and repeatedly over many periods; and (c) agents' optimal decisions include consideration of strategic complementarity. To avoid the curse of dimensionality that ordinarily prevents analysis of large networks, I introduce the idea of *network opacity* – that agents know who they observe, but not who anybody else observes. Instead, I suppose that agents know only the (common) distribution from which those observees are drawn.

This assumption grants that an arbitrarily accurate simulation may be performed by selecting a cut-off, k^*, on the number of higher-order expectations and a cut-off, p^*, on the number of compound expectations to consider. The first of these arises from the standard assumption that agents place decreasing weight on higher-order expectations. The second emerges from (a) the opacity of the network (so that agents are interested in a sequence of weighted-average expectations); (b) the recursive nature of the Kalman filter (so that each weighted-average expectation depends on the next-higher weighted average from the previous period); and (c) the AR process of the underlying state (so that older shocks are of decreasing importance to the current state).

Theorem 1 demonstrates that when the underlying state follows an VAR(1) process, the full hierarchy of relevant aggregate expectations will follow an VARMA(1,1) process with *network shocks* – weighted sums of agents' idiosyncratic shocks – entering both contemporaneously and with a lag.

A number of broad consequences of the model emerge directly from theorem 1. First, it is possible to simulate the effects of network learning without having to simulate the network explicitly: the network shocks together represent a sufficient statistic for the effect of the network on agents' aggregate beliefs. This makes the model particularly amenable to nesting within broad general equilibrium models of the economy.

Second, impulse responses of average expectations following shocks to the underlying state will exhibit greater persistence than the state itself, increasing in the number of agents observed. This is a form of rational herding behaviour that combines the herding exhibited in both Banerjee (1992), where agents observe others' actions, but have no strategic motive; and Morris and Shin (2002), where agents have a strategic motive, but do not observe others' actions.

Third, when the network is asymptotically irregular (i.e. has a distribution of links that is sufficiently far from uniform), mean zero idiosyncratic shocks do not wash out in aggregation, thereby leading to a network-based source of aggregate volatility, independent of "true" aggregate shocks to the hidden state. The scale of this additional volatility depends on the degree of irregularity in the network, which is captured simply in a single parameter: ζ^*.

Finally, because of the recursive nature of agents' learning, the aggregate effects of idiosyncratic shocks are *persistent*, even though the shocks themselves are entirely transitory.

The model would appear to be applicable to a variety of problems in macroeconomic research, including, for example, firms' price-setting decisions, labour search-and-matching models and asset pricing problems.

References

Acemoglu, D., V. Carvalho, A. Ozdaglar, and A. Tahbaz-Saleh (2012): "The Network Origins of Aggregate Fluctuations," *Econometrica*, 80(5), 1977–2016.

Acemoglu, D., M. Dahleh, I. Lobel, and A. Ozdaglar (2011): "Bayesian Learning in Social Networks," *Review of Economic Studies*, 78(4), 1201–1236.

Acemoglu, D., and A. Ozdaglar (2011): "Opinion Dynamics and Learning in Social Networks," *Dynamic Games and Applications*, 1, 3–49.

Albert, R., and A.-L. Barabasi (2002): "Statistical mechanics of complex networks," *Reviews of Modern Physics*, 74(1), 47–97.

Angeletos, G., and J. La'O (2009): "Incomplete Information, Higher-Order Beliefs and Price Inertia," *NBER Working Papers*, No. 15003.

——— (2010): "Noisy Business Cycles," in *NBER Macroeconomics Annual 2009*, ed. by D. Acemoglu, K. Rogoff, and M. Woodford, vol. 24, chap. 5, pp. 319–378. University of Chicago Press, Chicago.

Banerjee, A. (1992): "A simple model of herd benavior," *The Quarterly Journal of Economics*, 107(3), 797–817.

Banerjee, A., and D. Fudenberg (2004): "Word-of-mouth learning," *Games and Economic Behavior*, 46(1), 1–22.

Bierman, G. J. (1977): *Factorization Methods for Discrete Sequential Estimation*, Mathematics in Science and Engineering v. 128. Academic Press, New York.

Calvó-Armengol, A., and J. de Martí (2007): "Communication Networks: Knowledge and Decisions," *American Economic Review*, 97(2), 86–91.

Calvo, G. (1983): "Staggered Prices in a Utility-Maximizing Framework," *Journal of Monetary Economics*, 12(3), 383–398.

Clauset, A., C. R. Shalizi, and M. E. J. Newman (2009): "Power-Law Distributions in Empirical Data," *SIAM Review*, 51(4), 661–703.

DeGroot, M. (1974): "Reaching a Consensus," *Journal of the American Statistical Association*, 69(345), 118–121.

DeMarzo, P., D. Vayanos, and J. Zwiebel (2003): "Persuasion Bias, Social Influence and Unidimensional Opinions," *The Quarterly Journal of Economics*, 118(3), 909–968.

Evans, G. W., and S. Honkapohja (2001): *Learning and Expectations in Macroeconomics*. Princeton University Press, Princeton.

Gabaix, X. (2011): "The Granular Origins of Aggregate Volatility," *Econometrica*, 79(3), 733–772.

Gale, D., and S. Kariv (2003): "Bayesian learning in social networks," *Games and Economic Behavior*, 45(2), 329–346.

Golub, B., and M. Jackson (2010): "Naïve Learning in Social Networks and the Wisdom of Crowds," *American Economic Journal: Microeconomics*, 2(1), 112–149.

Goyal, S. (2007): *Connections: an introduction to the economics of networks*. Princeton University Press, Princeton.

Graham, L. (2011a): "Individual rationality, model-consistent expectations and learning," CDMA Working Paper Series 201112, Centre for Dynamic Macroeconomic Analysis.

——— (2011b): "Learning, information and heterogeneity," CDMA Working Paper Series 201113, Centre for Dynamic Macroeconomic Analysis.

Graham, L., and S. Wright (2010): "Information, hetereogeneity and market incompleteness," *Journal of Monetary Economics*, 57(2), 164–174.

Hamilton, J. (1994): *Time Series Analysis*. Princeton University Press, 1 edn.

Jackson, M. (2008): *Social and Economic Networks*. Princeton University Press, Princeton.

Jackson, M., and B. Rogers (2007): "Meeting Strangers and Friends of Friends: How Random Are Social Networks?," *American Economic Review*, 97(3), 890–915.

Kemeny, J., and J. L. Snell (1960): *Finite Markov Chains*. Van Nostrand, Princeton.

Keynes, J. M. (1936): *The General Theory of Employment, Interest and Money*. Palgrave Macmillan.

Kohlhas, A. (2013): "Learning-by-sharing: Monetary Policy and the Information Content of Aggregate Variables," Job market paper, Cambridge.

Lee, I. H. (1993): "On the Convergence of Informational Cascades," *Journal of Economic Theory*, 61(2), 395–411.

Ljungqvist, L., and T. Sargent (2004): *Recursive Macroeconomic Theory*. MIT Press, 2 edn.

Lorenzoni, G. (2009): "A Theory of Demand Shocks," *American Economic Review*, 99(5), 2055–2084.

Lucas, R. (1972): "Expectations and the Neutrality of Money," *Journal of Economic Theory*, 4(2), 103–124.

——— (1976): "Econometric Policy Evaluation: A Critique," in *The Phillips Curve and Labor Markets*, ed. by K. Brunner, and A. Meltzer, pp. 19–46. American Elsevier, New York.

Mankiw, G., and R. Reis (2002): "Sticky information versus sticky prices: A proposal to replace the New Keynsian Phillips curve," *The Quarterly Journal of Economics*, 117(4), 1295–1328.

——— (2006): "Pervasive Stickiness," *American Economic Review*, 96(2), 164–169.

——— (2007): "Sticky Information in General Equilibrium," *Journal of the European Economic Association*, 5(2-3), 603–613.

Melosi, L. (2014): "Estimating Models with Dispersed Information," *American Economic Journal: Macroeconomics*, 6, 1–31.

Morris, S., and H. S. Shin (2002): "Social Value of Public Information," *American Economic Review*, 92(5), 1521–1534.

Mueller-Frank, M. (2013): "A General Framework for Rational Learning in Social Networks," *Theoretical Economics*, Forthcoming(1), 1–40.

Nimark, K. (2008): "Dynamic pricing and imperfect common knowledge," *Journal of Monetary Economics*, 55(2), 365–382.

——— (2011a): "Dynamic Higher Order Expectations," Universitat Pompeu Fabra Economics Working Papers No 1118.

——— (2011b): "A low dimensional Kalman Filter for systems with lagged observables," mimeo.

Phelps, E. (1984): "The trouble with 'rational expectations' and the problem of inflation stabilization," in *Individual Forecasting and Aggregate Outcomes: 'Rational Expectations' Examined*, ed. by R. Frydman, and E. Phelps, chap. 2, pp. 31–46. Cambridge University Press.

Schennach, S. (2013): "Long memory via networking," CeMMAP working papers CWP13/13, Centre for Microdata Methods and Practice, Institute for Fiscal Studies, UCL.

Sims, C. (2003): "Implications of Rational Inattention," *Journal of Monetary Economics*, 50(3), 665–690.

Smith, L., and P. Sørensen (2000): "Pathological Outcomes of Observational Learning," *Econometrica*, 68(2), 371–398.

Thornton, C. L. (1976): "Triangular Covariance Factorizations for Kalman Filtering," Ph.D. thesis, University of California at Los Angeles.

Townsend, R. M. (1983): "Forecasting the Forecasts of Others," *Journal of Political Economy*, 91(4), 546–587.

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

Appendix: Peering into the mist: Social learning over an opaque observation network

John Barrdear*
Bank of England and the Centre for Macroeconomics

4th June 2014

1 Proof of Lemma 1.

I here demonstrate that agents' contemporaneous expectations of the network shocks are zero:

$$E_t(i)[\boldsymbol{z}_t] = \boldsymbol{0} \; \forall i, t \tag{1}$$

Since all shocks are Gaussian, the ability of an agent to create an expectation about a variable depends on the covariance between that variable and the agent's signal vector. But, by construction, agent i does not observe any signal that is based on $\boldsymbol{z}_t$. Since $\boldsymbol{z}_t$ is transitory and fully independent across time and from the underlying state, it must be the case that $Cov(\boldsymbol{z}_t, \boldsymbol{s}_t(i)) = \boldsymbol{0}$. The only possible exception to this is to note that $\boldsymbol{z}_t$ is comprised of weighted sums of idiosyncratic shocks and agent i's signals do include $\boldsymbol{v}_t(i)$. However, it must be that:

$$\begin{aligned} Cov\left({}^{1:}\widetilde{\boldsymbol{v}}_t, \boldsymbol{v}_t(i)\right) &= E\left[\lim_{n\to\infty}\sum_{j=1}^{n}\phi_n(j)\,\boldsymbol{v}_t(j)\,\boldsymbol{v}_t(i)\right] \\ &= \lim_{n\to\infty}\phi_n(i)\,\Sigma_{vv} \\ &= \boldsymbol{0} \end{aligned}$$

where the second equality relies on the independence of agents' idiosyncratic shocks and the third on assumption 2 (which grants us that $\lim_{n\to\infty}\phi_n(i) = 0 \; \forall i$). An equivalent argument applies to all higher-weighted averages: $Cov\left({}^{p:}\widetilde{\boldsymbol{v}}_t, \boldsymbol{v}_t(j)\right)$.

2 Proof of lemma 2.

The Kalman filter requires that each agent construct a prior expectation of the signal she will receive and then update her beliefs on the basis of the extent to which the signal she actually receives is a surprise. Using the equation for each agent's decision rule, we have that when preparing for period $t+1$, agent i will construct her prior expectation of her social signal as follows:

$$E_t(i)[g_t(\delta_t(i))] = E_t(i)\left[\boldsymbol{\lambda}_1' E_t(\delta_t(i))[X_t] + \boldsymbol{\lambda}_2'\boldsymbol{x}_t + \boldsymbol{\lambda}_3'\boldsymbol{v}_t(\delta_t(i))\right]$$

*Email: john.barrdear@bankofengland.co.uk

Recall that $\delta_t(i)$ is not known to agent i until period $t+1$. By denying agents knowledge of the full network and, instead, granting them knowledge of the distribution from which observation links are drawn ($\boldsymbol{\Phi}$) and using the assumption that this distribution is independent of other shocks, we can note that:

$$\begin{aligned}
E_t(i)\left[g_t(\delta_t(i))\right] &= \int_0^1 E_t(i)\left[g_t(j)\right]\phi(j)\,dj \\
&= E_t(i)\left[\int_0^1 g_t(j)\,\phi(j)\,dj\right] \\
&= E_t(i)\left[\widetilde{g}_t\right] \\
&= E_t(i)\left[\boldsymbol{\lambda}_1'\widetilde{E}_t\left[X_t\right] + \boldsymbol{\lambda}_2'\boldsymbol{x}_t + \boldsymbol{\lambda}_3'\widetilde{\boldsymbol{v}}_t\right]
\end{aligned}$$

where the second equality exploits the linearity of the expectation operator. The object $\widetilde{g}_t \equiv \int_0^1 g_t(j)\,\phi(j)\,dj$ is a *weighted* average of all agents' actions in period t using the observation p.d.f. as the weights.

3 Proof of proposition 1.

Denoting $\zeta(n) \equiv \sum_{i=1}^n \phi_n(i)^2$ and assuming that $\lim_{n\to\infty}\zeta(n) = \zeta^* \in (0,1)$ (assumption 2), we here demonstrate the following results regarding agents' idiosyncratic shocks:

1. ${}^{p:\sim}\widetilde{\boldsymbol{v}}_{n,t} \xrightarrow{d} {}^{p:\sim}\widetilde{\boldsymbol{v}}_t\ \forall p$ where ${}^{p:\sim}\widetilde{\boldsymbol{v}}_t \sim N\left(\boldsymbol{0}, \Sigma_{\utilde{vv}}^{\{p\}}\right) \quad \Sigma_{\utilde{vv}}^{\{q\}} = \left(1-(1-\zeta^*)^q\right)\Sigma_{vv}$
2. ${}^{p:\bullet\bullet}\ddot{\boldsymbol{v}}_{n,t} \xrightarrow{\mathcal{L}^2} {}^{p:\sim}\widetilde{\boldsymbol{v}}_t\ \forall p$
3. $Cov\left({}^{p:\sim}\widetilde{\boldsymbol{v}}_t, {}^{r:\sim}\widetilde{\boldsymbol{v}}_t\right) = \Sigma_{\utilde{vv}}^{\{p\}}\ \forall r < q$

where the weighted sums are defined as:

$$\begin{aligned}
{}^{1:\sim}\widetilde{\boldsymbol{v}}_{n,t} &\equiv \frac{1}{n}\sum_{i=1}^n \boldsymbol{v}_t(\delta_t(i)) & {}^{1:\bullet\bullet}\ddot{\boldsymbol{v}}_{n,t} &\equiv \sum_{i=1}^n \boldsymbol{v}_t(i)\,\phi_n(i) \\
{}^{2:\sim}\widetilde{\boldsymbol{v}}_{n,t} &\equiv \frac{1}{n}\sum_{i=1}^n \boldsymbol{v}_t(\delta_t(\delta_t(i))) & {}^{2:\bullet\bullet}\ddot{\boldsymbol{v}}_{n,t} &\equiv \sum_{i=1}^n \boldsymbol{v}_t(\delta_t(i))\,\phi_n(i) \\
{}^{3:\sim}\widetilde{\boldsymbol{v}}_{n,t} &\equiv \frac{1}{n}\sum_{i=1}^n \boldsymbol{v}_t(\delta_t(\delta_t(\delta_t(i)))) & {}^{3:\bullet\bullet}\ddot{\boldsymbol{v}}_{n,t} &\equiv \sum_{i=1}^n \boldsymbol{v}_t(\delta_t(\delta_t(i)))\,\phi_n(i) \\
&\vdots & &\vdots
\end{aligned}$$

First, note that since the vector $\boldsymbol{v}_t(i)$ is drawn from independent and identical Gaussian distributions with mean zero for each i and t, all of the weighted sums must also be distributed normally with mean zero. We now consider each of the results in turn.

1. ${}^{p:\sim}\widetilde{\boldsymbol{v}}_{n,t} \xrightarrow{d} {}^{p:\sim}\widetilde{\boldsymbol{v}}_t\ \forall p \quad {}^{p:\sim}\widetilde{\boldsymbol{v}}_t \sim N\left(\boldsymbol{0}, \Sigma_{\utilde{vv}}^{\{p\}}\right) \quad \Sigma_{\utilde{vv}}^{\{p\}} = \left(1-(1-\zeta^*)^p\right)\Sigma_{vv}$

Since it is clear that ${}^{p:\sim}\widetilde{\boldsymbol{v}}_{n,t}$ must converge to a normal distribution with mean zero, all that remains is to determine its variance-covariance matrix (note that the law of large numbers will apply here when the variance-covariance matrix is zero).

We will begin by considering each weighted-sum in turn.

- ${}^{1:}\tilde{\boldsymbol{v}}_{n,t} \xrightarrow{d} {}^{1:}\tilde{\boldsymbol{v}}_{t}$

The variance of ${}^{1:}\tilde{\boldsymbol{v}}_{n,t}$ is given by:

$$\begin{aligned}
Var\left[{}^{1:}\tilde{\boldsymbol{v}}_{n,t}\right] &= \frac{1}{n^2} Var\left[\boldsymbol{v}_t\left(\delta_t\left(1\right)\right) + \boldsymbol{v}_t\left(\delta_t\left(2\right)\right) + \cdots + \boldsymbol{v}_t\left(\delta_t\left(n\right)\right)\right] \\
&= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} E\left[\boldsymbol{v}_t\left(\delta_t\left(i\right)\right) \boldsymbol{v}_t\left(\delta_t\left(j\right)\right)\right] \\
&= \frac{1}{n^2} \left(n\Sigma_{vv} + \sum_{i=1}^{n} \sum_{j\neq i}^{n} E\left[\boldsymbol{v}_t\left(\delta_t\left(i\right)\right) \boldsymbol{v}_t\left(\delta_t\left(j\right)\right)\right] \right)
\end{aligned}$$

However, when $i \neq j$, given the full independence of the distributions of agents' observees, it must be that

$$\begin{aligned}
E\left[\boldsymbol{v}_t\left(\delta_t\left(i\right)\right) \boldsymbol{v}_t\left(\delta_t\left(j\right)\right)\right] &= \sum_{k=1}^{n} \phi_n\left(k\right) E\left[\boldsymbol{v}_t\left(k\right) \boldsymbol{v}_t\left(\delta_t\left(j\right)\right)\right] \\
&= \sum_{k=1}^{n} \phi_n\left(k\right) \left(\sum_{l=1}^{n} \phi_n\left(l\right) E\left[\boldsymbol{v}_t\left(k\right) \boldsymbol{v}_t\left(l\right)\right] \right) \\
&= \sum_{k=1}^{n} \phi_n\left(k\right)^2 E\left[\boldsymbol{v}_t\left(k\right) \boldsymbol{v}_t\left(k\right)\right] \\
&= \zeta\left(n\right) \Sigma_{vv}
\end{aligned} \tag{2}$$

where in moving from the second line to the third we have made use of the independence of agents' idiosyncratic shocks. We therefore have that

$$\begin{aligned}
Var\left[{}^{1:}\tilde{\boldsymbol{v}}_{n,t}\right] &= \frac{1}{n^2} \left(n\Sigma_{vv} + \sum_{i=1}^{n} \sum_{j\neq i}^{n} \zeta\left(n\right) \Sigma_{vv} \right) \\
&= \frac{1}{n^2} \left(n\Sigma_{vv} + \left(n^2 - n\right) \zeta\left(n\right) \Sigma_{vv} \right) \\
&= \frac{1}{n}\Sigma_{vv} + \left(\frac{n-1}{n}\right) \zeta\left(n\right) \Sigma_{vv}
\end{aligned}$$

and thus, in the limit, it must be that

$$\Sigma^{\{1\}}_{\underset{\sim}{v}\underset{\sim}{v}} \equiv \lim_{n\to\infty} Var\left[{}^{1:}\tilde{\boldsymbol{v}}_{n,t}\right] = \zeta^{*}\Sigma_{vv} \tag{3}$$

- ${}^{2:\sim}\tilde{\boldsymbol{v}}_{n,t} \xrightarrow{d} {}^{2:\sim}\tilde{\boldsymbol{v}}_{t}$

The variance of ${}^{2:\sim}\tilde{\boldsymbol{v}}_{n,t}$ is given by:

$$\begin{aligned}
Var\left[{}^{2:\sim}\tilde{\boldsymbol{v}}_{n,t}\right] &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} E\left[\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(i\right)\right)\right)\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(j\right)\right)\right)\right] \\
&= \frac{1}{n^2}\left(n\Sigma_{vv} + \sum_{i=1}^{n}\sum_{j\neq i}^{n} E\left[\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(i\right)\right)\right)\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(j\right)\right)\right)\right]\right)
\end{aligned}$$

Focussing on the latter term, we have that when $i \neq j$, it must be that

$$\begin{aligned}
E\left[\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(i\right)\right)\right)\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(j\right)\right)\right)\right] &= \sum_{k=1}^{n}\phi_n\left(k\right)E\left[\boldsymbol{v}_t\left(\delta_t\left(k\right)\right)\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(j\right)\right)\right)\right] \\
&= \sum_{k=1}^{n}\phi_n\left(k\right)\left(\sum_{l=1}^{n}\phi_n\left(l\right)E\left[\boldsymbol{v}_t\left(\delta_t\left(k\right)\right)\boldsymbol{v}_t\left(\delta_t\left(l\right)\right)\right]\right) \\
&= \sum_{k=1}^{n}\phi_n\left(k\right)^2\Sigma_{vv} \\
&\quad + \sum_{k=1}^{n}\sum_{l\neq k}^{n}\phi_n\left(k\right)\phi_n\left(l\right)E\left[\boldsymbol{v}_t\left(\delta_t\left(k\right)\right)\boldsymbol{v}_t\left(\delta_t\left(l\right)\right)\right]
\end{aligned}$$

It was shown above in equation (2) that

$$E\left[\boldsymbol{v}_t\left(\delta_t\left(k\right)\right)\boldsymbol{v}_t\left(\delta_t\left(l\right)\right)\right] = \zeta\left(n\right)\Sigma_{vv}\ \forall k \neq l$$

so it follows that

$$E\left[\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(i\right)\right)\right)\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(j\right)\right)\right)\right] = \zeta\left(n\right)\Sigma_{vv} + \zeta\left(n\right)\Sigma_{vv}\sum_{k=1}^{n}\sum_{l\neq k}^{n}\phi_n\left(k\right)\phi_n\left(l\right)$$

next, consider that since $\phi_n\left(k\right)$ and $\phi_n\left(l\right)$ are p.d.fs, it must be that

$$\begin{aligned}
\sum_{k=1}^{n}\sum_{l=1}^{n}\phi_n\left(i\right)\phi_n\left(j\right) &= \sum_{k=1}^{n}\phi_n\left(k\right)\left(\sum_{l=1}^{n}\phi_n\left(l\right)\right) \\
&= \sum_{k=1}^{n}\phi_n\left(k\right) \\
&= 1
\end{aligned}$$

We must therefore have that

$$\sum_{k=1}^{n}\sum_{l\neq k}^{n}\phi_n\left(k\right)\phi_n\left(l\right) = 1 - \sum_{k=1}^{n}\phi_n\left(k\right)^2 = 1 - \zeta\left(n\right) \tag{4}$$

Thus, when $i \neq j$, we have

$$E\left[\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(i\right)\right)\right)\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(j\right)\right)\right)\right] = \zeta\left(n\right)\Sigma_{vv} + \left(1-\zeta\left(n\right)\right)\zeta\left(n\right)\Sigma_{vv} \tag{5}$$

Substituting this back in, we arrive at

$$\begin{aligned}
Var\left[{}^{2:\sim}\tilde{\boldsymbol{v}}_{n,t}\right] &= \frac{1}{n}\Sigma_{vv} + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left(\zeta\left(n\right)\Sigma_{vv} + \left(1-\zeta\left(n\right)\right)\zeta\left(n\right)\Sigma_{vv}\right) \\
&= \frac{1}{n}\Sigma_{vv} + \frac{n\left(n-1\right)}{n^2}\left(\zeta\left(n\right)\Sigma_{vv} + \left(1-\zeta\left(n\right)\right)\zeta\left(n\right)\Sigma_{vv}\right)
\end{aligned}$$

and thus, in the limit, it must be that

$$\Sigma_{\underset{\sim}{\tilde{v}}\underset{\sim}{\tilde{v}}}^{\{2\}} \equiv \lim_{n\to\infty} Var\left[{}^{2:\sim}\tilde{\boldsymbol{v}}_{n,t}\right] = \zeta^*\Sigma_{vv} + \left(1-\zeta^*\right)\zeta^*\Sigma_{vv} \tag{6}$$

- $\overset{3:\sim}{\boldsymbol{v}}_{n,t} \xrightarrow{d} \overset{3:\sim}{\boldsymbol{v}}_t$

The variance of $\overset{3:\sim}{\boldsymbol{v}}_{n,t}$ is given by:

$$\begin{aligned} Var\left[\overset{3:\sim}{\boldsymbol{v}}_{n,t}\right] &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} E\left[\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(\delta_t\left(i\right)\right)\right)\right)\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(\delta_t\left(j\right)\right)\right)\right)\right] \\ &= \frac{1}{n^2}\left(n\Sigma_{vv} + \sum_{i=1}^{n}\sum_{j\neq i}^{n} E\left[\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(\delta_t\left(i\right)\right)\right)\right)\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(\delta_t\left(j\right)\right)\right)\right)\right]\right) \end{aligned}$$

Focussing on the latter term, we have that when $i \neq j$, it must be that

$$\begin{aligned} &E\left[\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(\delta_t\left(i\right)\right)\right)\right)\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(\delta_t\left(j\right)\right)\right)\right)\right] \\ &\quad= \sum_{k=1}^{n}\phi_n\left(k\right)\left(\sum_{l=1}^{n}\phi_n\left(l\right)E\left[\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(k\right)\right)\right)\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(l\right)\right)\right)\right]\right) \\ &\quad= \sum_{k=1}^{n}\phi_n\left(k\right)^2\Sigma_{vv} + \sum_{k=1}^{n}\sum_{l\neq k}^{n}\phi_n\left(k\right)\phi_n\left(l\right)E\left[\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(k\right)\right)\right)\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(l\right)\right)\right)\right] \end{aligned}$$

It was shown above in equation (5) that

$$E\left[\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(k\right)\right)\right)\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(l\right)\right)\right)\right] = \zeta\left(n\right)\Sigma_{vv} + \left(1-\zeta\left(n\right)\right)\zeta\left(n\right)\Sigma_{vv}$$

Combined with equation (4), this then implies that when $i \neq j$,

$$\begin{aligned} &E\left[\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(\delta_t\left(i\right)\right)\right)\right)\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(\delta_t\left(j\right)\right)\right)\right)\right] \\ &\quad= \zeta\left(n\right)\Sigma_{vv} + \left(1-\zeta\left(n\right)\right)\left(\zeta\left(n\right)\Sigma_{vv} + \left(1-\zeta\left(n\right)\right)\zeta\left(n\right)\Sigma_{vv}\right) \end{aligned} \tag{7}$$

Substituting this back in, we arrive at

$$\begin{aligned} Var\left[\overset{3:\sim}{\boldsymbol{v}}_{n,t}\right] &= \frac{1}{n}\Sigma_{vv} \\ &\quad+ \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left(\zeta\left(n\right)\Sigma_{vv} + \left(1-\zeta\left(n\right)\right)\left(\zeta\left(n\right)\Sigma_{vv} + \left(1-\zeta\left(n\right)\right)\zeta\left(n\right)\Sigma_{vv}\right)\right) \\ &= \frac{1}{n}\Sigma_{vv} \\ &\quad+ \frac{n\left(n-1\right)}{n^2}\left(\zeta\left(n\right)\Sigma_{vv} + \left(1-\zeta\left(n\right)\right)\left(\zeta\left(n\right)\Sigma_{vv} + \left(1-\zeta\left(n\right)\right)\zeta\left(n\right)\Sigma_{vv}\right)\right) \end{aligned}$$

and thus, in the limit, it must be that

$$\Sigma^{\{3\}}_{\underset{vv}{\sim\sim}} \equiv \lim_{n\to\infty} Var\left[\overset{3:\sim}{\boldsymbol{v}}_{n,t}\right] = \zeta^*\Sigma_{vv} + \left(1-\zeta^*\right)\left(\zeta^*\Sigma_{vv} + \left(1-\zeta^*\right)\zeta^*\Sigma_{vv}\right) \tag{8}$$

- The general case

By this stage, it should be clear that the variance-covariance matricies of higher weighted averages of agents' idiosyncratic shocks are able to be expressed in a recursive form:

$$\Sigma^{\{q\}}_{\underset{\sim}{vv}} = \zeta^* \Sigma_{vv} + (1-\zeta^*)\, \Sigma^{\{q-1\}}_{\underset{\sim}{vv}}$$

This may be simplified by first expanding it as

$$\begin{aligned}
\Sigma^{\{q\}}_{\underset{\sim}{vv}} &= \left(\sum_{p=0}^{q-1} (1-\zeta^*)^p \right) \zeta^* \Sigma_{vv} \\
&= \left(\frac{1-(1-\zeta^*)^q}{1-(1-\zeta^*)} \right) \zeta^* \Sigma_{vv} \\
&= \left(1-(1-\zeta^*)^q\right) \Sigma_{vv} \qquad (9)
\end{aligned}$$

which completes the proof of the first result.

As a matter of curiosity, this result also obtains from the following when taking each variable in $\boldsymbol{v}_t\,(i)$ separately (for simplicity I have shown only three agents, when there are actually $n \to \infty$):

$$\Sigma^{\{1\}}_{\underset{\sim}{vv}} = \phi' \begin{bmatrix} \sigma_v^2 & 0 & 0 \\ 0 & \sigma_v^2 & 0 \\ 0 & 0 & \sigma_v^2 \end{bmatrix} \phi = \phi'\phi\, \sigma_v^2$$

$$\begin{aligned}
\Sigma^{\{2\}}_{\underset{\sim}{vv}} &= \phi' \begin{bmatrix} \sigma_v^2 & \Sigma^{\{1\}}_{\underset{\sim}{vv}} & \Sigma^{\{1\}}_{\underset{\sim}{vv}} \\ \Sigma^{\{1\}}_{\underset{\sim}{vv}} & \sigma_v^2 & \Sigma^{\{1\}}_{\underset{\sim}{vv}} \\ \Sigma^{\{1\}}_{\underset{\sim}{vv}} & \Sigma^{\{1\}}_{\underset{\sim}{vv}} & \sigma_v^2 \end{bmatrix} \phi = \phi' \begin{bmatrix} \sigma_v^2 & \phi' \begin{bmatrix} \sigma_v^2 & 0 & 0 \\ 0 & \sigma_v^2 & 0 \\ 0 & 0 & \sigma_v^2 \end{bmatrix} \phi & \phi' \begin{bmatrix} \sigma_v^2 & 0 & 0 \\ 0 & \sigma_v^2 & 0 \\ 0 & 0 & \sigma_v^2 \end{bmatrix} \phi \\ \phi' \begin{bmatrix} \sigma_v^2 & 0 & 0 \\ 0 & \sigma_v^2 & 0 \\ 0 & 0 & \sigma_v^2 \end{bmatrix} \phi & \sigma_v^2 & \phi' \begin{bmatrix} \sigma_v^2 & 0 & 0 \\ 0 & \sigma_v^2 & 0 \\ 0 & 0 & \sigma_v^2 \end{bmatrix} \phi \\ \phi' \begin{bmatrix} \sigma_v^2 & 0 & 0 \\ 0 & \sigma_v^2 & 0 \\ 0 & 0 & \sigma_v^2 \end{bmatrix} \phi & \phi' \begin{bmatrix} \sigma_v^2 & 0 & 0 \\ 0 & \sigma_v^2 & 0 \\ 0 & 0 & \sigma_v^2 \end{bmatrix} \phi & \sigma_v^2 \end{bmatrix} \phi \\
&= \phi' \begin{bmatrix} 1 & \phi'\phi & \phi'\phi \\ \phi'\phi & 1 & \phi'\phi \\ \phi'\phi & \phi'\phi & 1 \end{bmatrix} \phi\, \sigma_v^2 \\
&= \phi'\phi \left(1 + (1-\phi'\phi)\right) \sigma_v^2
\end{aligned}$$

$$\Sigma^{\{p\}}_{\underset{\sim}{vv}} = \phi' \begin{bmatrix} \sigma_v^2 & \Sigma^{\{p-1\}}_{\underset{\sim}{vv}} & \Sigma^{\{p-1\}}_{\underset{\sim}{vv}} \\ \Sigma^{\{p-1\}}_{\underset{\sim}{vv}} & \sigma_v^2 & \Sigma^{\{p-1\}}_{\underset{\sim}{vv}} \\ \Sigma^{\{p-1\}}_{\underset{\sim}{vv}} & \Sigma^{\{p-1\}}_{\underset{\sim}{vv}} & \sigma_v^2 \end{bmatrix} \phi = \phi'\phi \left(1 + (1-\phi'\phi) + \cdots + (1-\phi'\phi)^{p-1}\right) \sigma_v^2$$

2. ${}^{p:}\ddot{\boldsymbol{v}}_{n,t} \xrightarrow{\mathcal{L}^2} {}^{p:}\tilde{\boldsymbol{v}}_t \; \forall q$

We next demonstrate that ${}^{p:}\ddot{\boldsymbol{v}}_{n,t}$ converges to ${}^{p:}\tilde{\boldsymbol{v}}_t$ in mean square error.[1] That is, we show that $\lim_{n\to\infty} E\left[\left({}^{p:}\ddot{\boldsymbol{v}}_{n,t} - {}^{p:}\tilde{\boldsymbol{v}}_t\right)^2\right] = 0$. First, see that:

$$
\begin{aligned}
E\left[\left({}^{p:}\ddot{\boldsymbol{v}}_{n,t} - {}^{2:}\tilde{\boldsymbol{v}}_t\right)^2\right] &= E\left[\left({}^{p:}\ddot{\boldsymbol{v}}_{n,t}\right)^2 - 2\,{}^{p:}\ddot{\boldsymbol{v}}_{n,t}\tilde{\boldsymbol{v}}_t + \left({}^{p:}\tilde{\boldsymbol{v}}_t\right)^2\right] \\
&= Var\left[{}^{p:}\ddot{\boldsymbol{v}}_{n,t}\right] - 2Cov\left[{}^{p:}\ddot{\boldsymbol{v}}_{n,t}, {}^{p:}\tilde{\boldsymbol{v}}_t\right] + Var\left[{}^{p:}\tilde{\boldsymbol{v}}_t\right]
\end{aligned}
$$

The third term is just $\Sigma^{\{q\}}_{\tilde{v}\tilde{v}}$ from the first result above. We now consider the first and second terms in turn. The variance of ${}^{p:}\ddot{\boldsymbol{v}}_{n,t}$ is given by:

$$
\begin{aligned}
Var\left[{}^{p:}\ddot{\boldsymbol{v}}_{n,t}\right] &= Var\left[\sum_{i=1}^{n} \phi_n(i)\, \boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t}_{p-1}(i)))\right)\right] \\
&= E\left[\sum_{i=1}^{n}\sum_{j=1}^{n} \phi_n(i)\,\phi_n(j)\, \boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t}_{p-1}(i)))\right) \boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t}_{p-1}(j)))\right)\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \phi_n(i)\,\phi_n(j)\, E\left[\boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t}_{p-1}(i)))\right) \boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t}_{p-1}(j)))\right)\right] \\
&= \sum_{i=1}^{n} \phi_n(i)^2\, \Sigma_{vv} \\
&\quad + \sum_{i=1}^{n}\sum_{j\neq i}^{n} \phi_n(i)\,\phi_n(j)\, E\left[\boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t}_{p-1}(i)))\right) \boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t}_{p-1}(j)))\right)\right]
\end{aligned}
$$

But we know from the first result above that when $i \neq j$,

$$
\begin{aligned}
& E\left[\boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t}_{p-1}(i)))\right) \boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t}_{p-1}(j)))\right)\right] \\
&= \zeta(n)\,\Sigma_{vv} + (1-\zeta(n))\, E\left[\boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t}_{p-2}(i)))\right) \boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t}_{p-2}(j)))\right)\right]
\end{aligned}
$$

noting the recursive structure and making use of equation (4) then gives us

$$
Var\left[{}^{p:}\ddot{\boldsymbol{v}}_{n,t}\right] = \zeta(n)\,\Sigma_{vv} + (1-\zeta(n))\, Var\left[{}^{p-1:}\ddot{\boldsymbol{v}}_{n,t}\right]
$$

which, in the limit, becomes

$$
\lim_{n\to\infty} Var\left[{}^{p:}\ddot{\boldsymbol{v}}_{n,t}\right] = \zeta^*\Sigma_{vv} + (1-\zeta^*) \lim_{n\to\infty} Var\left[{}^{p-1:}\ddot{\boldsymbol{v}}_{n,t}\right]
$$

which is the same rule for $Var\left[{}^{p:}\tilde{\boldsymbol{v}}_{n,t}\right]$, which implies that

$$
\lim_{n\to\infty} Var\left[{}^{p:}\ddot{\boldsymbol{v}}_{n,t}\right] = \lim_{n\to\infty} Var\left[{}^{p:}\tilde{\boldsymbol{v}}_{n,t}\right] = \Sigma^{\{p\}}_{\tilde{v}\tilde{v}}
$$

[1] Recall that convergence in mean square error is a stronger form of convergence than convergence in probability.

Turning next to the covariance between ${}^{p:\cdot\cdot}\boldsymbol{v}_{n,t}$ and ${}^{p:\sim}\boldsymbol{v}_t$, we note that

$$
\begin{aligned}
Cov\left[{}^{p:\cdot\cdot}\boldsymbol{v}_{n,t},\ {}^{p:\sim}\boldsymbol{v}_{n,t}\right] &= E\left[\begin{array}{l}\left(\sum_{i=1}^{n}\phi_n(i)\,\boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t}_{p-1}(i)))\right)\right)\\ \times\left(\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t}_{p}(j)))\right)\right)\end{array}\right]\\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\phi_n(i)\,E\left[\boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t}_{p-1}(i)))\right)\boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t}_{p}(j)))\right)\right]\\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\phi_n(i)\,\phi_n(k)\,E\left[\begin{array}{l}\boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t}_{p-1}(i)))\right)\\ \times\boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t}_{p-1}(k)))\right)\end{array}\right]
\end{aligned}
$$

where moving from the second line to the third makes use of the independence of agents' draws from Φ_n and the linearity of the expectation operator. This, in turn, may be rewritten as

$$
Cov\left[{}^{p:\cdot\cdot}\boldsymbol{v}_{n,t},\ {}^{p:\sim}\boldsymbol{v}_{n,t}\right] = \frac{n}{n}\left(\begin{array}{l}\sum_{i=1}^{n}\phi_n(i)^2\,\Sigma_{vv}\\ +\sum_{i=1}^{n}\sum_{k\neq i}^{n}\phi_n(i)\,\phi_n(k)\,E\left[\begin{array}{l}\boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t}_{p-1}(i)))\right)\\ \times\boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t}_{p-1}(k)))\right)\end{array}\right]\end{array}\right)
$$

Since this is the same expression as that for $Var\left[{}^{p:\cdot\cdot}\boldsymbol{v}_{n,t}\right]$ above, we therefore have

$$
\lim_{n\to\infty} Cov\left[{}^{p:\cdot\cdot}\boldsymbol{v}_{n,t},\ {}^{p:\sim}\boldsymbol{v}_{n,t}\right] = \Sigma_{\widetilde{vv}}^{\{p\}}
$$

and, hence, that

$$
\begin{aligned}
\lim_{n\to\infty} E\left[\left({}^{p:\cdot\cdot}\boldsymbol{v}_{n,t} - {}^{p:\sim}\boldsymbol{v}_t\right)^2\right] &= \Sigma_{\widetilde{vv}}^{\{p\}} - 2\Sigma_{\widetilde{vv}}^{\{p\}} + \Sigma_{\widetilde{vv}}^{\{p\}}\\
&= 0
\end{aligned}
$$

as required.

3. $Cov\left[{}^{p:\sim}\boldsymbol{v}_t,\ {}^{r:\sim}\boldsymbol{v}_t\right] = \Sigma_{\widetilde{vv}}^{\{p\}}\ \forall p < r$

To prove this, we will first consider $Cov\left[{}^{p:\sim}\boldsymbol{v}_t,\ {}^{p+1:\sim}\boldsymbol{v}_t\right]$ and later consider $r \geq p+2$.

$$
\begin{aligned}
Cov\left[{}^{\{p\}}\widetilde{\boldsymbol{v}}_{n,t},\ {}^{\{p+1\}}\widetilde{\boldsymbol{v}}_{n,t}\right] &= E\left[\begin{array}{l}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{v}_t\left(\underbrace{\delta_t\cdots\delta_t}_{p}(i)\right)\right)\\ \times\left(\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{v}_t\left(\underbrace{\delta_t\cdots\delta_t}_{p+1}(j)\right)\right)\end{array}\right]\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left[\boldsymbol{v}_t\left(\underbrace{\delta_t\cdots\delta_t}_{p}(i)\right)\boldsymbol{v}_t\left(\underbrace{\delta_t\cdots\delta_t}_{p+1}(j)\right)\right]
\end{aligned}
$$

Focussing on the final term, note that

$$
\begin{aligned}
&E\left[\boldsymbol{v}_t\left(\underbrace{\delta_t\cdots\delta_t}_{p}(i)\right)\boldsymbol{v}_t\left(\underbrace{\delta_t\cdots\delta_t}_{p+1}(j)\right)\right] \\
&=\sum_{k=1}^{n}\phi_n(k)\,E\left[\boldsymbol{v}_t\left(\underbrace{\delta_t\cdots\delta_t}_{p}(i)\right)\boldsymbol{v}_t\left(\underbrace{\delta_t\cdots\delta_t}_{p}(k)\right)\right] \\
&=\phi_n(i)\,\Sigma_{vv}+\sum_{k\neq i}^{n}\phi_n(k)\,E\left[\boldsymbol{v}_t\left(\underbrace{\delta_t\cdots\delta_t}_{p}(i)\right)\boldsymbol{v}_t\left(\underbrace{\delta_t\cdots\delta_t}_{p}(k)\right)\right] \\
&=\phi_n(i)\,\Sigma_{vv}+(1-\phi_n(i))\,\Sigma^{p}_{\underset{\sim}{v}\underset{\sim}{v}}(n)
\end{aligned}
$$

Substituting this back into the above then gives

$$
\begin{aligned}
Cov\left[{}^{p:\sim}\boldsymbol{v}_t,{}^{p+1:\sim}\boldsymbol{v}_t\right]&=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\phi_n(i)\,\Sigma_{vv}+(1-\phi_n(i))\,\Sigma^{p}_{\underset{\sim}{v}\underset{\sim}{v}}(n)\right) \\
&=\frac{1}{n}\sum_{i=1}^{n}\left(\phi_n(i)\,\Sigma_{vv}+(1-\phi_n(i))\,\Sigma^{p}_{\underset{\sim}{v}\underset{\sim}{v}}(n)\right) \\
&=\frac{1}{n}\Sigma_{vv}+\frac{1}{n}\sum_{i=1}^{n}(1-\phi_n(i))\,\Sigma^{p}_{\underset{\sim}{v}\underset{\sim}{v}}(n)
\end{aligned}
$$

In the limit, this becomes

$$
\lim_{n\to\infty}Cov\left[{}^{p:\sim}\boldsymbol{v}_{n,t},{}^{p+1:\sim}\boldsymbol{v}_{n,t}\right]=\Sigma^{p}_{\underset{\sim}{v}\underset{\sim}{v}}
$$

which establishes the result for $r=p+1$. For $r=p+2$, note that

$$
\begin{aligned}
&E\left[\boldsymbol{v}_t\left(\underbrace{\delta_t\cdots\delta_t}_{p}(i)\right)\boldsymbol{v}_t\left(\underbrace{\delta_t\cdots\delta_t}_{p+2}(j)\right)\right] \\
&=\sum_{k=1}^{n}\phi_n(k)\,E\left[\boldsymbol{v}_t\left(\underbrace{\delta_t\cdots\delta_t}_{p}(i)\right)\boldsymbol{v}_t\left(\underbrace{\delta_t\cdots\delta_t}_{p+1}(k)\right)\right] \\
&=\sum_{k=1}^{n}\sum_{l=1}^{n}\phi_n(k)\,\phi_n(l)\,E\left[\boldsymbol{v}_t\left(\underbrace{\delta_t\cdots\delta_t}_{p}(i)\right)\boldsymbol{v}_t\left(\underbrace{\delta_t\cdots\delta_t}_{p}(l)\right)\right] \\
&=\sum_{l=1}^{n}\phi_n(l)\,E\left[\boldsymbol{v}_t\left(\underbrace{\delta_t\cdots\delta_t}_{p}(i)\right)\boldsymbol{v}_t\left(\underbrace{\delta_t\cdots\delta_t}_{p}(l)\right)\right]
\end{aligned}
$$

which is the same as for $r=p+1$. It should be clear that this same process would apply for all $r\geq p+2$, which establishes the result.

4 Proof of theorem 1.

The state vector of interest and its law of motion are conjectured to be:

$$X_t \equiv \begin{bmatrix} \boldsymbol{x}_t \\ \overline{E}_t [X_t] \\ \overset{1:\sim}{E}_t [X_t] \\ \overset{2:\sim}{E}_t [X_t] \\ \vdots \end{bmatrix} = F X_{t-1} + G_1 \boldsymbol{u}_t + G_2 \boldsymbol{z}_t + G_3 \boldsymbol{e}_t + G_4 \boldsymbol{z}_{t-1} \tag{10}$$

while agents' private/public and social signals are given by:

$$\boldsymbol{s}_t^p (i) = D_1 \boldsymbol{x}_t + D_2 X_{t-1} + R_1 \boldsymbol{v}_t (i) + R_2 \boldsymbol{e}_t + R_3 \boldsymbol{z}_{t-1} \tag{11a}$$

$$\boldsymbol{s}_t^s (i) = \boldsymbol{\lambda}_1' E_{t-1} (\delta_{t-1} (i)) [X_{t-1}] + \boldsymbol{\lambda}_2' \boldsymbol{x}_{t-1} + \boldsymbol{\lambda}_3' \boldsymbol{v}_{t-1} (\delta_{t-1} (i)) \tag{11b}$$

Together, these describe a linear state space system to which a Kalman filter provides the optimal linear estimator (in the sense of minimising mean squared error).

As discussed in the main text, the system described here is not in the form of a classic state space problem, both because of the presence of the lagged state in agents' signals and because of the moving average component of the law of motion. Lemma 1 demonstrated that we do not need to include $\boldsymbol{z}_t$ in the agents' state vector of interest. To deal with the lagged observations, we follow Nimark (2011b) in developing a modified Kalman filter that does not require the stacking of the state vectors of interest.

To begin, we define the matrices S_x, T_s and T_{w_p} as the matrices that select $\boldsymbol{x}_t$, $\overline{E}_t [X_t]$ and $\overset{p:\sim}{E}_t [X_t]$ respectively from X_t (e.g., $T_{w_2} X_t = \overset{2:\sim}{E}_t [X_t]$).

We also define the general notation that $\theta_{t|q}^{\text{err}} (i)$ represents the error in agent i's period-q expectation regarding θ_t. In particular, we will use the following:

$$\begin{aligned} \boldsymbol{s}_{t|t-1}^{\text{err}} (i) &\equiv \boldsymbol{s}_t (i) - E_{t-1} (i) [\boldsymbol{s}_t (i)] && : \text{signal innovation} \\ X_{t|t-1}^{\text{err}} (i) &\equiv X_t - E_{t-1} (i) [X_t] && : \text{prior error} \\ X_{t|t}^{\text{err}} (i) &\equiv X_t - E_t (i) [X_t] && : \text{contemporaneous error} \end{aligned}$$

4.1 The filter

We proceed by deploying a Gram-Schmidt orthogonalisation of agents' signals. That is, noting that the signal innovation

$$\boldsymbol{s}_{t|t-1}^{\text{err}} (i) \equiv \boldsymbol{s}_t (i) - E_{t-1} (i) [\boldsymbol{s}_t (i)] \tag{12}$$

contains only *new* information available to i in period t, we conclude that it must be orthogonal to any of j's estimates based on information from earlier periods. We can therefore use the standard result that $E[x|y,z] = E[x|y] + E[x|z]$ when $y \perp z$, so that

$$\begin{aligned} E_t (i) [X_t] &= E [X_t | \mathcal{I}_{t-1} (i)] + E \left[X_t | \boldsymbol{s}_{t|t-1}^{\text{err}} (i) \right] \\ &= E_{t-1} (i) [X_t] + K_t \boldsymbol{s}_{t|t-1}^{\text{err}} (i) \end{aligned} \tag{13}$$

for some projection matrix, K_t (the Kalman gain). Note that K_t does not require an agent subscript as the problem is symmetric for all agents.

Optimality then requires that the projection matrix, K_t, be such that the signal innovation, $\boldsymbol{s}^{\text{err}}_{t|t-1}(i)$, is orthogonal to the projection error, $X_t - K_t \boldsymbol{s}^{\text{err}}_{t|t-1}(i)$. That is, we require that

$$E\left[\left(X_t - K_t \boldsymbol{s}^{\text{err}}_{t|t-1}(i)\right) \boldsymbol{s}^{\text{err}}_{t|t-1}(i)'\right] = 0$$

Rearranging then gives an expression for the optimal Kalman gain:

$$K_t = E\left[X_t \boldsymbol{s}^{\text{err}}_{t|t-1}(i)'\right] \left(E\left[\boldsymbol{s}^{\text{err}}_{t|t-1}(i)\, \boldsymbol{s}^{\text{err}}_{t|t-1}(i)'\right]\right)^{-1} \quad \forall i \tag{14}$$

which, since the unconditional expectations of X_t and all signal innovations are zero, is simply

$$K_t = Cov(X_t, \boldsymbol{s}^{\text{err}}_{t|t-1}(i)) \left[Var\left(\boldsymbol{s}^{\text{err}}_{t|t-1}(i)\right)\right]^{-1}$$

In order to evaluate this, it is necessary to construct expressions for the innovation in agents' private and social signals. We consider each in turn.

Agents' private signals

To begin, we substitute the conjectured state law of motion into the private signal equation to get:

$$\begin{aligned} \boldsymbol{s}^p_t(j) = &(D_1 S_x F + D_2) X_{t-1} + D_1 S_x G_1 \boldsymbol{u}_t \\ &+ R_1 \boldsymbol{v}_t(j) + R_2 \boldsymbol{e}_t + R_3 \boldsymbol{z}_{t-1} \end{aligned} \tag{15}$$

where we have used the fact that $\boldsymbol{x}_t$ is independent of network shocks to ignore the $G_2 \boldsymbol{z}_t$ and $G_4 \boldsymbol{z}_{t-1}$ components of X_t. From this, we see that i's prior expectation of her private signal will be given by

$$E_{t-1}(i)\left[\boldsymbol{s}^p_t(i)\right] = (D_1 S_x F + D_2) E_{t-1}(i)\left[X_{t-1}\right] \tag{16}$$

where we have made use of lemma 1 to drop the term in $E_{t-1}(i)\left[\boldsymbol{z}_{t-1}\right]$. Subtracting equation (16) from (15) then gives the innovation in agents' private signals as

$$\begin{aligned} \boldsymbol{s}^p_{t|t-1}(i) = &(D_1 S_x F + D_2) X^{\text{err}}_{t-1|t-1}(i) + D_1 S_x G_1 \boldsymbol{u}_t \\ &+ R_1 \boldsymbol{v}_t(j) + R_2 \boldsymbol{e}_t + R_3 \boldsymbol{z}_{t-1} \end{aligned} \tag{17}$$

where $X^{\text{err}}_{t|t}(i)$ is i's contemporaneous error in estimating X_t.

Agents' social signals

For the social signal, and assuming temporarily that agents observe the actions of only one competitor, we make use of proposition 1 to write the prior expectation as

$$E_{t-1}(i)\left[\boldsymbol{s}^s_t(i)\right] = \boldsymbol{\lambda}_1' E_{t-1}(i)\left[\widetilde{E}_{t-1}\left[X_{t-1}\right]\right] + \boldsymbol{\lambda}_2' E_{t-1}(i)\left[\boldsymbol{x}_{t-1}\right] + \boldsymbol{\lambda}_3' E_{t-1}(i)\left[\widetilde{\boldsymbol{v}}_{t-1}\right]$$

Given that $E_t(i)\left[\boldsymbol{z}_t\right] = 0$, $S_x X_t = \boldsymbol{x}_t$ and $T_{w_1} X_t = {}^{\{1\}}\widetilde{E}_t\left[X_t\right]$, we can write this as

$$E_{t-1}(i)\left[\boldsymbol{s}^s_t(i)\right] = \left(\boldsymbol{\lambda}_2' S_x + \boldsymbol{\lambda}_1' T_w\right) E_{t-1}(i)\left[X_{t-1}\right] \tag{18}$$

Subtracting (18) from (11b), we then have that the innovation in the agent's social signal is given by:

$$\begin{aligned} \boldsymbol{s}^s_{t|t-1}(i) = &\boldsymbol{\lambda}_2' S_x X^{\text{err}}_{t-1|t-1}(i) \\ &+ \boldsymbol{\lambda}_1' E_{t-1}(\delta_{t-1}(i))\left[X_{t-1}\right] - \boldsymbol{\lambda}_1' T_w E_{t-1}(i)\left[X_{t-1}\right] \\ &+ \boldsymbol{\lambda}_3' \boldsymbol{v}_{t-1}(\delta_{t-1}(i)) \end{aligned}$$

Adding and subtracting $\boldsymbol{\lambda}_1' T_w X_{t-1}$ on the right-hand side then gives

$$\begin{aligned}\boldsymbol{s}^s_{t|t-1}(i) &= \left(\boldsymbol{\lambda}_2' S_x + \boldsymbol{\lambda}_1' T_w\right) X^{\text{err}}_{t-1|t-1}(i) \\ &\quad - \boldsymbol{\lambda}_1'\left(T_w X_{t-1} - E_{t-1}\left(\delta_{t-1}(i)\right)\left[X_{t-1}\right]\right) \\ &\quad + \boldsymbol{\lambda}_3' \boldsymbol{v}_{t-1}\left(\delta_{t-1}(i)\right)\end{aligned}$$

and finally now adding and subtracting $\boldsymbol{\lambda}_1' X_{t-1}$ on the right-hand side gives

$$\begin{aligned}\boldsymbol{s}^s_{t|t-1}(i) &= \left(\boldsymbol{\lambda}_2' S_x + \boldsymbol{\lambda}_1' T_w\right) X^{\text{err}}_{t-1|t-1}(i) - \boldsymbol{\lambda}_1' X^{\text{err}}_{t-1|t-1}\left(\delta_{t-1}(i)\right) \\ &\quad + \boldsymbol{\lambda}_1'\left(I - T_w\right) X_{t-1} \\ &\quad + \boldsymbol{\lambda}_3' \boldsymbol{v}_{t-1}\left(\delta_{t-1}(i)\right)\end{aligned}$$

Crucially, we have that the innovation in i's social signal includes not only a term in their own contemporaneous error from the previous period but also a term in their *observee*'s error.

The combined signal innovation

Stacking the private, public and social signal innovations, we then obtain

$$\begin{aligned}\boldsymbol{s}^{\text{err}}_{t|t-1}(i) &= M_1 X^{\text{err}}_{t-1|t-1}(i) + M_2 X^{\text{err}}_{t-1|t-1}\left(\delta_{t-1}(i)\right) + M_3 X_{t-1} \\ &\quad + N_1 \boldsymbol{u}_t + N_2 \boldsymbol{v}_t(i) + N_3 \boldsymbol{e}_t + N_4 \boldsymbol{v}_{t-1}\left(\delta_{t-1}(i)\right) + N_5 \boldsymbol{z}_{t-1}\end{aligned} \tag{19a}$$

where

$$M_1 = \begin{bmatrix} D_1 S_x F + D_2 \\ \boldsymbol{\lambda}_2' S_x + \boldsymbol{\lambda}_1' T_w \end{bmatrix} \quad M_2 = \begin{bmatrix} \mathbf{0} \\ -\boldsymbol{\lambda}_1' \end{bmatrix} \quad M_3 = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\lambda}_1'\left(I - T_w\right) \end{bmatrix} \tag{19b}$$

$$N_1 = \begin{bmatrix} D_1 S_x G_1 \\ \mathbf{0} \end{bmatrix} \quad N_2 = \begin{bmatrix} R_1 \\ \mathbf{0} \end{bmatrix} \quad N_3 = \begin{bmatrix} R_2 \\ \mathbf{0} \end{bmatrix} \quad N_4 = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\lambda}_3' \end{bmatrix} \quad N_5 = \begin{bmatrix} R_3 \\ \mathbf{0} \end{bmatrix} \tag{19c}$$

Considering two or more observees is then obtained by further stacking the signals

$$\begin{aligned}\boldsymbol{s}^{\text{err}}_{t|t-1}(i) &= M_1 X^{\text{err}}_{t-1|t-1}(i) + M_2 \begin{bmatrix} X^{\text{err}}_{t-1|t-1}\left(\delta_{t-1}(i,1)\right) \\ X^{\text{err}}_{t-1|t-1}\left(\delta_{t-1}(i,2)\right) \end{bmatrix} + M_3 X_{t-1} \\ &\quad + N_1 \boldsymbol{u}_t + N_2 \boldsymbol{v}_t(i) + N_3 \boldsymbol{e}_t + N_4 \begin{bmatrix} \boldsymbol{v}_{t-1}\left(\delta_{t-1}(i,1)\right) \\ \boldsymbol{v}_{t-1}\left(\delta_{t-1}(i,2)\right) \end{bmatrix} + N_5 \boldsymbol{z}_{t-1}\end{aligned} \tag{20a}$$

where

$$M_1 = \begin{bmatrix} D_1 S_x F + D_2 \\ \boldsymbol{\lambda}_2' S_x + \boldsymbol{\lambda}_1' T_w \\ \boldsymbol{\lambda}_2' S_x + \boldsymbol{\lambda}_1' T_w \end{bmatrix} \quad M_2 = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ -\boldsymbol{\lambda}_1' & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{\lambda}_1' \end{bmatrix} \quad M_3 = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\lambda}_1'\left(I - T_w\right) \\ \boldsymbol{\lambda}_1'\left(I - T_w\right) \end{bmatrix} \tag{20b}$$

$$N_1 = \begin{bmatrix} D_1 S_x G_1 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \quad N_2 = \begin{bmatrix} R_1 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \quad N_3 = \begin{bmatrix} R_2 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \quad N_4 = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \boldsymbol{\lambda}_3' & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\lambda}_3' \end{bmatrix} \quad N_5 = \begin{bmatrix} R_3 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \tag{20c}$$

For the remainder of this appendix, we shall use the notation of a single observee on the understanding that the signal innovation may be replace as above for an arbitrary number of competitors observed.

Deriving the Kalman gain

We first expand the first term in equation (14) as

$$E\left[X_t \boldsymbol{s}^{\text{err}}_{t|t-1}(i)'\right] = E\left[\begin{array}{l} \left(FX_{t-1} + G_1\boldsymbol{u}_t + G_2\boldsymbol{z}_t + G_4\boldsymbol{z}_{t-1} + G_3\boldsymbol{e}_t\right) \\ \times \left(\begin{array}{l} M_1 X^{\text{err}}_{t-1|t-1}(i) \\ +M_2 X^{\text{err}}_{t-1|t-1}(\delta_{t-1}(i)) \\ +M_3 X_{t-1} \\ +N_1\boldsymbol{u}_t + N_2\boldsymbol{v}_t(i) + N_3\boldsymbol{e}_t \\ +N_4\boldsymbol{v}_{t-1}(\delta_{t-1}(i)) + N_5\boldsymbol{z}_{t-1} \end{array}\right)' \end{array}\right]$$

$$= E\left[\begin{array}{l} (FX_{t-1})\left(M_1 X^{\text{err}}_{t-1|t-1}(i)\right)' \\ +(FX_{t-1})\left(M_2 X^{\text{err}}_{t-1|t-1}(\delta_{t-1}(i))\right)' \\ +(FX_{t-1})(M_3X_{t-1})' \\ +(FX_{t-1})(N_5\boldsymbol{z}_{t-1})' \\ +(G_1\boldsymbol{u}_t)(N_1\boldsymbol{u}_t)' \\ +(G_3\boldsymbol{e}_t)(N_3\boldsymbol{e}_t)' \\ +(G_4\boldsymbol{z}_{t-1})\left(M_1 X^{\text{err}}_{t-1|t-1}(i)\right)' \\ +(G_4\boldsymbol{z}_{t-1})\left(M_2 X^{\text{err}}_{t-1|t-1}(\delta_{t-1}(i))\right)' \\ +(G_4\boldsymbol{z}_{t-1})(M_3X_{t-1})' \\ +(G_4\boldsymbol{z}_{t-1})(N_5\boldsymbol{z}_{t-1})' \end{array}\right] \tag{21}$$

where we use the fact that period-t shocks are orthogonal to period-$(t-1)$ objects and make use of assumption 2 (which grants us that $\lim_{n\to\infty}\phi_n(i) = 0\ \forall i$) to note that there is no covariance between period-$(t-1)$ objects and $\boldsymbol{v}_{t-1}(i)\ \forall i$.

next, we note that for any j and any t, we may write

$$\begin{aligned} E\left[X_t X^{\text{err}}_{t|t}(j)'\right] &= E\left[\left(X^{\text{err}}_{t|t}(j) + E_t(j)[X_t]\right) X^{\text{err}}_{t|t}(j)'\right] \\ &= E\left[X^{\text{err}}_{t|t}(j)\, X^{\text{err}}_{t|t}(j)'\right] \\ &= V_{t|t} \end{aligned}$$

where the second equality makes use of the fact that since $E_t(j)[X_t]$ is spanned by the set of orthogonal signal innovations $\left\{\boldsymbol{s}^{\text{err}}_{t|t-1}(j), \boldsymbol{s}^{\text{err}}_{t-1|t-2}(j), \cdots\right\}$ and these are orthogonal to $X^{\text{err}}_{t|t}(j)$ by construction, then it must be that $E_t(j)[X_t]$ and $X^{\text{err}}_{t|t}(j)$ are orthogonal for all j and t. Note that $V_{t|t} \equiv E\left[X^{\text{err}}_{t|t}(j)\, X^{\text{err}}_{t|t}(j)'\right]\ \forall j$ is the variance of each agent's contemporaneous error (common to all agents as their problems are symmetric).

Using this, we may rewrite (21) as

$$\begin{aligned} E\left[X_t \boldsymbol{s}^{\text{err}}_{t|t-1}(i)'\right] &= FV_{t-1|t-1}M_1' \\ &+ FV_{t-1|t-1}M_2' \\ &+ FU_{t-1}M_3' \\ &+ FG_2\Sigma_{zz}N_5' \\ &+ G_1\Sigma_{uu}N_1' \\ &+ G_3\Sigma_{ee}N_3' \\ &+ G_4\Sigma_{zz}G_2'\left(M_1 + M_2 + M_3\right)' \\ &+ G_4\Sigma_{zz}N_5' \end{aligned}$$

or, defining $M \equiv \begin{bmatrix} M_1 & M_2 & M_3 \end{bmatrix}$, as simply

$$
\begin{aligned}
E\left[X_t \boldsymbol{s}_{t|t-1}^{\text{err}}(i)'\right] = {} & F\begin{bmatrix} V_{t-1|t-1} & V_{t-1|t-1} & U_{t-1} \end{bmatrix} M' \\
& + G_1 \Sigma_{uu} N_1' \\
& + F G_2 \Sigma_{zz} N_5' \\
& + G_3 \Sigma_{ee} N_3' \\
& + G_4 \Sigma_{zz} G_2' \left(M_1 + M_2 + M_3\right)' \\
& + G_4 \Sigma_{zz} N_5'
\end{aligned} \tag{22}
$$

Turning to the second term in equation (14), we have that

$$
\begin{aligned}
E\left[\boldsymbol{s}_{t|t-1}^{\text{err}}(i)\, \boldsymbol{s}_{t|t-1}^{\text{err}}(i)'\right] &= E\left[\begin{array}{c}
\left(\begin{array}{l}
M_1 X_{t-1|t-1}^{\text{err}}(i) \\
+M_2 X_{t-1|t-1}^{\text{err}}(\delta_{t-1}(i)) \\
+M_3 X_{t-1} \\
+N_1 \boldsymbol{u}_t + N_2 \boldsymbol{v}_t(i) \\
+N_4 \boldsymbol{v}_{t-1}(\delta_{t-1}(i)) + N_5 \boldsymbol{z}_{t-1} + N_3 \boldsymbol{e}_t
\end{array}\right) \\
\times \left(\begin{array}{l}
M_1 X_{t-1|t-1}^{\text{err}}(i) \\
+M_2 X_{t-1|t-1}^{\text{err}}(\delta_{t-1}(i)) \\
+M_3 X_{t-1} \\
+N_1 \boldsymbol{u}_t + N_2 \boldsymbol{v}_t(i) \\
+N_4 \boldsymbol{v}_{t-1}(\delta_{t-1}(i)) + N_5 \boldsymbol{z}_{t-1} + N_3 \boldsymbol{e}_t
\end{array}\right)'
\end{array}\right] \\
&= E\left[\begin{array}{c}
\left(\begin{array}{l}
M_1 X_{t-1|t-1}^{\text{err}}(i) \\
+M_2 X_{t-1|t-1}^{\text{err}}(\delta_{t-1}(i)) \\
+M_3 X_{t-1} \\
+N_5 \boldsymbol{z}_{t-1}
\end{array}\right) \\
\times \left(\begin{array}{l}
M_1 X_{t-1|t-1}^{\text{err}}(i) \\
+M_2 X_{t-1|t-1}^{\text{err}}(\delta_{t-1}(i)) \\
+M_3 X_{t-1} \\
+N_5 \boldsymbol{z}_{t-1}
\end{array}\right)'
\end{array}\right] \\
&\quad + M_2 E\left[X_{t-1|t-1}^{\text{err}}(\delta_{t-1}(i))\, \boldsymbol{v}_{t-1}(\delta_{t-1}(i))'\right] N_4' \\
&\quad + N_4 E\left[\boldsymbol{v}_{t-1}(\delta_{t-1}(i))\, X_{t-1|t-1}^{\text{err}}(\delta_{t-1}(i))'\right] M_2' \\
&\quad + N_1 \Sigma_{uu} N_1' + N_2 \Sigma_{vv} N_2' + N_4 \Sigma_{vv} N_4' + N_3 \Sigma_{ee} N_3'
\end{aligned}
$$

Expanding out the various cross-products then gives us

$$\begin{aligned}E\left[\boldsymbol{s}^{\text{err}}_{t|t-1}(i)\,\boldsymbol{s}^{\text{err}}_{t|t-1}(i)'\right] &= M_1 V_{t-1|t-1} M_1' + M_1 W_{t-1|t-1} M_2' + M_1 V_{t-1|t-1} M_3' \\ &+ M_2 W_{t-1|t-1} M_1' + M_2 V_{t-1|t-1} M_2' + M_2 V_{t-1|t-1} M_3' \\ &+ M_3 V_{t-1|t-1} M_1' + M_3 V_{t-1|t-1} M_2' + M_3 U_{t-1} M_3' \\ &- M_2 K_{t-1} N_2 \Sigma_{vv} N_4' \\ &- N_4 \Sigma_{vv} N_2' K_{t-1}' M_2' \\ &+ N_1 \Sigma_{uu} N_1' + N_2 \Sigma_{vv} N_2' + N_4 \Sigma_{vv} N_4' \\ &+ (M_1 + M_2 + M_3)\, G_2 \Sigma_{zz} N_5' \\ &+ N_5 \Sigma_{zz} G_2' (M_1 + M_2 + M_3)' \\ &+ N_3 \Sigma_{ee} N_3'\end{aligned}$$

where $W_{t|t} \equiv E\left[X^{\text{err}}_{t|t}(i)\,X^{\text{err}}_{t|t}(j)'\right]\ \forall i \neq j$ is the covariance between any two agents' contemporaneous errors (common to all agent-pairs as their problems are symmetric and the network is opaque so they each have the same probability of observing the same target). Similarly to the covariance term, this may be written simply as

$$\begin{aligned}E\left[\boldsymbol{s}^{\text{err}}_{t|t-1}(i)\,\boldsymbol{s}^{\text{err}}_{t|t-1}(i)'\right] &= M \begin{bmatrix} V_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \\ W_{t-1|t-1} & V_{t-1|t-1} & V_{t-1|t-1} \\ V_{t-1|t-1} & V_{t-1|t-1} & U_{t-1} \end{bmatrix} M' \\ &- M_2 K_{t-1} N_2 \Sigma_{vv} N_4' \\ &- N_4 \Sigma_{vv} N_2' K_{t-1}' M_2' \\ &+ N_1 \Sigma_{uu} N_1' + N_2 \Sigma_{vv} N_2' + N_4 \Sigma_{vv} N_4' \\ &+ (M_1 + M_2 + M_3)\, G_2 \Sigma_{zz} N_5' \\ &+ N_5 \Sigma_{zz} G_2' (M_1 + M_2 + M_3)' \\ &+ N_3 \Sigma_{ee} N_3' \end{aligned} \tag{23}$$

Substituting (22) and (23) into (14) and gathering like terms, we arrive at:

$$K_t = \begin{pmatrix} F \begin{bmatrix} V_{t-1|t-1} & V_{t-1|t-1} & U_{t-1} \end{bmatrix} M' \\ +G_1 \Sigma_{uu} N_1' \\ +F G_2 \Sigma_{zz} N_5' \\ +G_4 \Sigma_{zz} G_2' (M_1 + M_2 + M_3)' \\ +G_4 \Sigma_{zz} N_5' \\ +G_3 \Sigma_{ee} N_3' \end{pmatrix} \times \begin{bmatrix} M \begin{bmatrix} V_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \\ W_{t-1|t-1} & V_{t-1|t-1} & V_{t-1|t-1} \\ V_{t-1|t-1} & V_{t-1|t-1} & U_{t-1} \end{bmatrix} M' \\ +(M_1 + M_2 + M_3)\, G_2 \Sigma_{zz} N_5' \\ +N_5 \Sigma_{zz} G_2' (M_1 + M_2 + M_3)' \\ -M_2 K_{t-1} N_2 \Sigma_{vv} N_4' \\ -N_4 \Sigma_{vv} N_2' K_{t-1}' M_2' \\ +N_1 \Sigma_{uu} N_1' + N_2 \Sigma_{vv} N_2' + N_4 \Sigma_{vv} N_4' + N_3 \Sigma_{ee} N_3' \end{bmatrix}^{-1} \tag{24}$$

4.2 Evolution of the variance-covariance matricies

Unconditional variance of the state vector of interest

From the conjectured law of motion, we can read immediately that the variance of the state vector of interest evolves as:

$$\begin{aligned} U_t &= FU_{t-1}F' \\ &+ G_1\Sigma_{uu}G_1' + G_2\Sigma_{zz}G_2' + G_3\Sigma_{ee}G_3' + G_4\Sigma_{zz}G_4' + FG_2\Sigma_{zz}G_4' + G_4\Sigma_{zz}G_2'F' \end{aligned} \tag{25}$$

Variance of agents' expectation errors

First, subtracting $E_{t-1}(i)[X_t]$ from each side of the state equation, we have:

$$\begin{aligned} X_t - E_{t-1}(i)[X_t] &= F(X_{t-1} - E_{t-1}(i)[X_{t-1}]) \\ &+ G_1\boldsymbol{u}_t + G_2\boldsymbol{z}_t + G_3\boldsymbol{e}_t + G_4\boldsymbol{z}_{t-1} \end{aligned} \tag{26}$$

Taking the variance of each side, we have that the prior variance will be given by:

$$\begin{aligned} V_{t|t-1} &= FV_{t-1|t-1}F' \\ &+ G_1\Sigma_{uu}G_1' + G_2\Sigma_{zz}G_2' + G_3\Sigma_{ee}G_3' + G_4\Sigma_{zz}G_4' + FG_2\Sigma_{zz}G_4' + G_4\Sigma_{zz}G_2'F' \end{aligned} \tag{27}$$

next, we subtract each side of equation (13) from X_t and rearrange to obtain

$$(X_t - E_t(i)[X_t]) + K_t\boldsymbol{s}^{\text{err}}_{t|t-1}(i) = (X_t - E_{t-1}(i)[X_t]) \tag{28}$$

Since the signal innovation is orthogonal to the contemporaneous error, $X_t - E_t(i)[X_t]$ by construction, the variance of the right-hand side must equal the sum of the variances on the left-hand side, thereby giving:

$$V_{t|t} + K_t\,Var\left(\boldsymbol{s}^{\text{err}}_{t|t-1}(i)\right)K_t' = V_{t|t-1}$$

or

$$V_{t|t} = V_{t|t-1} - K_t \left(\begin{array}{l} M \begin{bmatrix} V_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \\ W_{t-1|t-1} & V_{t-1|t-1} & V_{t-1|t-1} \\ V_{t-1|t-1} & V_{t-1|t-1} & U_{t-1} \end{bmatrix} M' \\ +(M_1+M_2+M_3)G_2\Sigma_{zz}N_5' \\ +N_5\Sigma_{zz}G_2'(M_1+M_2+M_3)' \\ -M_2K_{t-1}N_2\Sigma_{vv}N_4' \\ -N_4\Sigma_{vv}N_2'K_{t-1}'M_2' \\ +N_1\Sigma_{uu}N_1' + N_2\Sigma_{vv}N_2' + N_3\Sigma_{ee}N_3' + N_4\Sigma_{vv}N_4' \end{array} \right) K_t' \tag{29}$$

Covariance between agents' expectation errors

First, from (26), we have that the prior covariance between two agents' errors is given by:

$$\begin{aligned} W_{t|t-1} &\equiv E\left[X^{\text{err}}_{t|t-1}(i)\,X_{t|t-1}(j)'\right] \\ &= FW_{t-1|t-1}F' \\ &+ G_1\Sigma_{uu}G_1' + G_2\Sigma_{zz}G_2' + G_3\Sigma_{ee}G_3' + G_4\Sigma_{zz}G_4' + FG_2\Sigma_{zz}G_4' + G_4\Sigma_{zz}G_2'F' \end{aligned} \tag{30}$$

next, returning to equation (28)

$$\left(X_t - E_t\left(i\right)\left[X_t\right]\right) = \left(X_t - E_{t-1}\left(i\right)\left[X_t\right]\right) - K_t \boldsymbol{s}^{\text{err}}_{t|t-1}\left(i\right) \tag{31}$$

note that agent i's signal innovation will not necessarily be orthogonal to either of j's expectation errors, so we expand this fully to obtain

$$\begin{aligned} W_{t|t} &= W_{t|t-1} \\ &\quad + K_t\, Cov\left(\boldsymbol{s}^{\text{err}}_{t|t-1}\left(i\right), \boldsymbol{s}^{\text{err}}_{t|t-1}\left(j\right)\right) K_t' \\ &\quad - Cov\left(X^{\text{err}}_{t|t-1}\left(i\right), \boldsymbol{s}^{\text{err}}_{t|t-1}\left(j\right)\right) K_t' \\ &\quad - K_t\, Cov\left(\boldsymbol{s}^{\text{err}}_{t|t-1}\left(i\right), X_{t|t-1}\left(j\right)\right) \end{aligned} \tag{32}$$

For the second term on the right-hand side, we have

$$\begin{aligned} E\left[\boldsymbol{s}^{\text{err}}_{t|t-1}\left(i\right) \boldsymbol{s}^{\text{err}}_{t|t-1}\left(j\right)'\right] &= E\left[\begin{array}{c} \begin{pmatrix} M_1 X^{\text{err}}_{t-1|t-1}\left(i\right) \\ +M_2 X^{\text{err}}_{t-1|t-1}\left(\delta_{t-1}\left(i\right)\right) \\ +M_3 X_{t-1} \\ +N_1 \boldsymbol{u}_t + N_2 \boldsymbol{v}_t\left(i\right) \\ +N_4 \boldsymbol{v}_{t-1}\left(\delta_{t-1}\left(i\right)\right) + N_5 \boldsymbol{z}_{t-1} + N_3 \boldsymbol{e}_t \end{pmatrix} \\ \times \begin{pmatrix} M_1 X_{t-1|t-1}\left(j\right) \\ +M_2 X_{t-1|t-1}\left(\delta_{t-1}\left(j\right)\right) \\ +M_3 X_{t-1} \\ +N_1 \boldsymbol{u}_t + N_2 \boldsymbol{v}_t\left(j\right) \\ +N_4 \boldsymbol{v}_{t-1}\left(\delta_{t-1}\left(j\right)\right) + N_5 \boldsymbol{z}_{t-1} + N_3 \boldsymbol{e}_t \end{pmatrix}' \end{array}\right] \\ &= E\left[\begin{array}{c} \begin{pmatrix} M_1 X^{\text{err}}_{t-1|t-1}\left(i\right) \\ +M_2 X^{\text{err}}_{t-1|t-1}\left(\delta_{t-1}\left(i\right)\right) \\ +M_3 X_{t-1} \\ +N_5 \boldsymbol{z}_{t-1} \end{pmatrix} \\ \times \begin{pmatrix} M_1 X_{t-1|t-1}\left(j\right) \\ +M_2 X_{t-1|t-1}\left(\delta_{t-1}\left(j\right)\right) \\ +M_3 X_{t-1} \\ +N_5 \boldsymbol{z}_{t-1} \end{pmatrix}' \end{array}\right] \\ &\quad + N_1 \Sigma_{uu} N_1' \\ &\quad + N_3 \Sigma_{ee} N_3' \end{aligned}$$

Given $i \neq j$ and assumption 2, it must be the case that i, j, $\delta_{t-1}\left(i\right)$ and $\delta_{t-1}\left(j\right)$ are four different agents, almost surely. We therefore have

$$\begin{aligned} E\left[\boldsymbol{s}^{\text{err}}_{t|t-1}\left(i\right) \boldsymbol{s}^{\text{err}}_{t|t-1}\left(j\right)'\right] &= M \begin{bmatrix} W_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \\ W_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \\ V_{t-1|t-1} & V_{t-1|t-1} & U_{t-1} \end{bmatrix} M' \\ &\quad + \left(M_1 + M_2 + M_3\right) G_2 \Sigma_{zz} N_5' \\ &\quad + N_5 \Sigma_{zz} G_2' \left(M_1 + M_2 + M_3\right)' \\ &\quad + N_1 \Sigma_{uu} N_1' \\ &\quad + N_3 \Sigma_{ee} N_3' \end{aligned} \tag{33}$$

For the third term, we have

$$
\begin{aligned}
Cov\left(X_{t|t-1}^{\text{err}}\left(i\right),\boldsymbol{s}_{t|t-1}^{\text{err}}\left(j\right)\right) &= E\left[\begin{pmatrix} FX_{t-1|t-1}\left(j\right) \\ +G_1\boldsymbol{u}_t \\ +G_2\boldsymbol{z}_t \\ +G_4\boldsymbol{z}_{t-1} \\ +G_3\boldsymbol{e}_t \end{pmatrix} \times \begin{pmatrix} M_1X_{t-1|t-1}^{\text{err}}\left(i\right) \\ +M_2X_{t-1|t-1}^{\text{err}}\left(\delta_{t-1}\left(i\right)\right) \\ +M_3X_{t-1} \\ +N_1\boldsymbol{u}_t+N_2\boldsymbol{v}_t\left(i\right) \\ +N_4\boldsymbol{v}_{t-1}\left(\delta_{t-1}\left(i\right)\right)+N_5\boldsymbol{z}_{t-1}+N_3\boldsymbol{e}_t \end{pmatrix}'\right] \\
&= F\begin{bmatrix} V_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \end{bmatrix} M' \\
&+ G_1\Sigma_{uu}N_1' \\
&+ FG_2\Sigma_{zz}N_5' \\
&+ G_4\Sigma_{zz}G_2'\left(M_1+M_2+M_3\right)' \\
&+ G_4\Sigma_{zz}N_5' \\
&+ G_3\Sigma_{zz}N_3'
\end{aligned} \tag{34}
$$

while the fourth term is the transpose of the same.

Filter summary

In summary, the filter evolves through the following system of equations:

$$
\begin{aligned}
E\left[\boldsymbol{s}_{t|t-1}^{\text{err}}\left(i\right)\boldsymbol{s}_{t|t-1}^{\text{err}}\left(i\right)'\right] &= M\begin{bmatrix} V_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \\ W_{t-1|t-1} & V_{t-1|t-1} & V_{t-1|t-1} \\ V_{t-1|t-1} & V_{t-1|t-1} & U_{t-1} \end{bmatrix} M' \\
&+ \left(M_1+M_2+M_3\right)G_2\Sigma_{zz}N_5' \\
&+ N_5\Sigma_{zz}G_2'\left(M_1+M_2+M_3\right)' \\
&- M_2K_{t-1}N_2\Sigma_{vv}N_4' \\
&- N_4\Sigma_{vv}N_2'K_{t-1}'M_2' \\
&+ N_1\Sigma_{uu}N_1'+N_2\Sigma_{vv}N_2'+N_4\Sigma_{vv}N_4'
\end{aligned} \tag{35a}
$$

$$
\begin{aligned}
E\left[\boldsymbol{s}_{t|t-1}^{\text{err}}\left(i\right)\boldsymbol{s}_{t|t-1}^{\text{err}}\left(j\right)'\right] &= M\begin{bmatrix} W_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \\ W_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \\ V_{t-1|t-1} & V_{t-1|t-1} & U_{t-1} \end{bmatrix} M' \\
&+ \left(M_1+M_2+M_3\right)G_2\Sigma_{zz}N_5' \\
&+ N_5\Sigma_{zz}G_2'\left(M_1+M_2+M_3\right)' \\
&+ N_1\Sigma_{uu}N_1'
\end{aligned} \tag{35b}
$$

$$
\begin{aligned}
E\left[X_t \boldsymbol{s}^{\text{err}}_{t|t-1}(i)'\right] = F &\begin{bmatrix} V_{t-1|t-1} & V_{t-1|t-1} & U_{t-1} \end{bmatrix} M' \\
&+ G_1 \Sigma_{uu} N_1' \\
&+ F G_2 \Sigma_{zz} N_5' \\
&+ G_4 \Sigma_{zz} G_2' \left(M_1 + M_2 + M_3\right)' \\
&+ G_4 \Sigma_{zz} N_5'
\end{aligned} \tag{35c}
$$

$$
\begin{aligned}
E\left[X^{\text{err}}_{t|t-1}(i)\, \boldsymbol{s}^{\text{err}}_{t|t-1}(j)'\right] = F &\begin{bmatrix} V_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \end{bmatrix} M' \\
&+ G_1 \Sigma_{uu} N_1' \\
&+ F G_2 \Sigma_{zz} N_5' \\
&+ G_4 \Sigma_{zz} G_2' \left(M_1 + M_2 + M_3\right)' \\
&+ G_4 \Sigma_{zz} N_5'
\end{aligned} \tag{35d}
$$

$$
K_t = E\left[X_t \boldsymbol{s}^{\text{err}}_{t|t-1}(i)'\right] \left(E\left[\boldsymbol{s}^{\text{err}}_{t|t-1}(i)\, \boldsymbol{s}^{\text{err}}_{t|t-1}(i)'\right]\right)^{-1} \tag{35e}
$$

$$
\begin{aligned}
U_t = {} & F U_{t-1} F' \\
& + G_1 \Sigma_{uu} G_1' + G_2 \Sigma_{zz} G_2' + G_4 \Sigma_{zz} G_4' + F G_2 \Sigma_{zz} G_4' + G_4 \Sigma_{zz} G_2' F'
\end{aligned} \tag{35f}
$$

$$
\begin{aligned}
V_{t|t-1} = {} & F V_{t-1|t-1} F' \\
& + G_1 \Sigma_{uu} G_1' + G_2 \Sigma_{zz} G_2' + G_4 \Sigma_{zz} G_4' + F G_2 \Sigma_{zz} G_4' + G_4 \Sigma_{zz} G_2' F'
\end{aligned} \tag{35g}
$$

$$
\begin{aligned}
W_{t|t-1} = {} & F W_{t-1|t-1} F' \\
& + G_1 \Sigma_{uu} G_1' + G_2 \Sigma_{zz} G_2' + G_4 \Sigma_{zz} G_4' + F G_2 \Sigma_{zz} G_4' + G_4 \Sigma_{zz} G_2' F'
\end{aligned} \tag{35h}
$$

$$
V_{t|t} = V_{t|t-1} - K_t\, E\left[\boldsymbol{s}^{\text{err}}_{t|t-1}(i)\, \boldsymbol{s}^{\text{err}}_{t|t-1}(i)'\right] K_t' \tag{35i}
$$

$$
\begin{aligned}
W_{t|t} = W_{t|t-1} &+ K_t E\left[\boldsymbol{s}^{\text{err}}_{t|t-1}(i)\, \boldsymbol{s}^{\text{err}}_{t|t-1}(j)'\right] K_t' \\
&- E\left[X^{\text{err}}_{t|t-1}(i)\, \boldsymbol{s}^{\text{err}}_{t|t-1}(j)'\right] K_t' \\
&- K_t E\left[\boldsymbol{s}^{\text{err}}_{t|t-1}(i)\, X_{t|t-1}(j)'\right]
\end{aligned} \tag{35j}
$$

Provided that all eigenvalues of F are within the unit circle, then there will exist a steady state (i.e. time-invariant) filter, found by iterating these equations forward until convergence is achieved.

4.3 Confirming the conjectured law of motion

The state vector of interest and its law of motion are conjectured to be:

$$X_t \equiv \begin{bmatrix} \boldsymbol{x}_t \\ \overline{E}_t [X_t] \\ \overset{1:\sim}{E}_t [X_t] \\ \overset{2:\sim}{E}_t [X_t] \\ \vdots \end{bmatrix} = F X_{t-1} + G_1 \boldsymbol{u}_t + G_2 \boldsymbol{z}_t + G_3 \boldsymbol{e}_t + G_4 \boldsymbol{z}_{t-1} \tag{36}$$

To confirm this law of motion, we first combining equations (13) and (20) to write the agents' filter as:

$$\begin{aligned} E_t (i) [X_t] = {} & F E_{t-1} (i) [X_{t-1}] \\ & + K \begin{pmatrix} M_1 (X_{t-1} - E_{t-1} (i) [X_{t-1}]) \\ +M_2 (X_{t-1} - E_{t-1} (\delta_{t-1} (i)) [X_{t-1}]) \\ +M_3 X_{t-1} \\ +N_1 \boldsymbol{u}_t + N_2 \boldsymbol{v}_t (i) + N_3 \boldsymbol{e}_t \\ +N_4 \boldsymbol{v}_{t-1} (\delta_{t-1} (i)) + N_5 \boldsymbol{z}_{t-1} \end{pmatrix} \end{aligned}$$

Gathering like terms gives

$$\begin{aligned} E_t (i) [X_t] = {} & K (M_1 + M_2 + M_3) X_{t-1} \\ & + (F - K M_1) E_{t-1} (i) [X_{t-1}] \\ & - K M_2 E_{t-1} (\delta_{t-1} (i)) [X_{t-1}] \\ & + K N_1 \boldsymbol{u}_t \\ & + K N_2 \boldsymbol{v}_t (i) \\ & + K N_3 \boldsymbol{e}_t \\ & + K N_4 \boldsymbol{v}_{t-1} (\delta_{t-1} (i)) \\ & + K N_5 \boldsymbol{z}_{t-1} \end{aligned} \tag{37}$$

Taking the simple average of equation (37) gives

$$\begin{aligned} \overline{E}_t [X_t] = {} & K (M_1 + M_2 + M_3) X_{t-1} \\ & + (F - K M_1) \overline{E}_{t-1} [X_{t-1}] \\ & - K M_2 \overset{1:\sim}{E}_{t-1} [X_{t-1}] \\ & + K N_1 \boldsymbol{u}_t \\ & + K N_3 \boldsymbol{e}_t \\ & + K N_4 \overset{1:\sim}{\boldsymbol{v}}_{t-1} \\ & + K N_5 \boldsymbol{z}_{t-1} \end{aligned}$$

where I have used proposition 1 to replace $\int_0^1 \boldsymbol{v}_{t-1} (\delta_{t-1} (i)) \, di$ with $\overset{1:\sim}{\boldsymbol{v}}_{t-1}$. But since $\overset{1:\sim}{\boldsymbol{v}}_{t-1}$ is part of $\boldsymbol{z}_{t-1}$, while $\overline{E}_{t-1} [X_{t-1}]$ and $\overset{1:\sim}{E}_{t-1} [X_{t-1}]$ are part of X_{t-1}, we can simplify this down to:

$$\begin{aligned} \overline{E}_t [X_t] = {} & \{ K (M_1 + M_2 + M_3) + (F - K M_1) T_s - K M_2 T_{w_1} \} X_{t-1} \\ & + K N_1 \boldsymbol{u}_t \\ & + K N_3 \boldsymbol{e}_t \\ & + K \left(\begin{bmatrix} N_4 & \mathbf{0}_{1 \times \infty} \end{bmatrix} + N_5 \right) \boldsymbol{z}_{t-1} \end{aligned} \tag{38}$$

next, taking the p-th weighted average of equation (37) gives

$$\begin{aligned}\overset{1:\sim}{E}_t[X_t] = &K(M_1+M_2+M_3)X_{t-1}\\ &+(F-KM_1)\overset{1:\sim}{E}_{t-1}[X_{t-1}]\\ &-KM_2\overset{p+1:\sim}{E}_{t-1}[X_{t-1}]\\ &+KN_1\boldsymbol{u}_t\\ &+KN_2\overset{p:\sim}{\boldsymbol{v}}_t\\ &+KN_3\boldsymbol{e}_t\\ &+KN_4\overset{p+1:\sim}{\boldsymbol{v}}_{t-1}\\ &+KN_5\boldsymbol{z}_{t-1}\end{aligned}$$

where the last two terms have again made use of proposition 1. From this, we read immediately that

$$\begin{aligned}\overset{p:\sim}{E}_t[X_t] = &\left\{K(M_1+M_2+M_3)+(F-KM_1)T_{w_p}-KM_2T_{w_{p+1}}\right\}X_{t-1}\\ &+KN_1\boldsymbol{u}_t\\ &+K\begin{bmatrix}\mathbf{0}_{1\times r(q-1)} & N_2 & \mathbf{0}_{1\times\infty}\end{bmatrix}\boldsymbol{z}_t\\ &+KN_3\boldsymbol{e}_t\\ &+K\left(\begin{bmatrix}\mathbf{0}_{1\times rq} & N_4 & \mathbf{0}_{1\times\infty}\end{bmatrix}+N_5\right)\boldsymbol{z}_{t-1}\end{aligned} \tag{39}$$

where r is the number of elements in each agents' vector of idiosyncratic shocks, $\boldsymbol{v}_t(i)$. Putting it all together, we substitute equations (38) and (39) into equation (36) to arrive at

$$F=\begin{bmatrix}\begin{bmatrix}A & \mathbf{0}_{m\times\infty}\end{bmatrix}\\ K(M_1+M_2+M_3)+(F-KM_1)T_s-KM_2T_{w_1}\\ K(M_1+M_2+M_3)+(F-KM_1)T_{w_1}-KM_2T_{w_2}\\ K(M_1+M_2+M_3)+(F-KM_1)T_{w_2}-KM_2T_{w_3}\\ \vdots\end{bmatrix} \tag{40a}$$

$$G_1=\begin{bmatrix}P\\ KN_1\\ KN_1\\ KN_1\\ \vdots\end{bmatrix}\qquad G_2=\begin{bmatrix}\mathbf{0}_{m\times\infty}\\ \mathbf{0}_{\infty\times\infty}\\ K\begin{bmatrix}N_2 & \mathbf{0}_{1\times r} & \mathbf{0}_{1\times r} & \mathbf{0}_{1\times\infty}\end{bmatrix}\\ K\begin{bmatrix}\mathbf{0}_{1\times r} & N_2 & \mathbf{0}_{1\times r} & \mathbf{0}_{1\times\infty}\end{bmatrix}\\ \vdots\end{bmatrix} \tag{40b}$$

$$G_3=\begin{bmatrix}\mathbf{0}_{m\times n}\\ KN_3\\ KN_3\\ KN_3\\ \vdots\end{bmatrix}\qquad G_4=\begin{bmatrix}\mathbf{0}_{m\times\infty}\\ K\left(\begin{bmatrix}N_4 & \mathbf{0}_{1\times p} & \mathbf{0}_{1\times r} & \mathbf{0}_{1\times\infty}\end{bmatrix}+N_5\right)\\ K\left(\begin{bmatrix}\mathbf{0}_{1\times p} & N_4 & \mathbf{0}_{1\times r} & \mathbf{0}_{1\times\infty}\end{bmatrix}+N_5\right)\\ K\left(\begin{bmatrix}\mathbf{0}_{1\times r} & \mathbf{0}_{1\times r} & N_4 & \mathbf{0}_{1\times\infty}\end{bmatrix}+N_5\right)\\ \vdots\end{bmatrix} \tag{40c}$$

where m is the number of elements in the underlying state $(\boldsymbol{x}_t)$ and n is the number of elements in the vector of public signal noise $(\boldsymbol{e}_t)$. This confirms the conjectured structure to the law of motion and implicitly defines the coefficient matricies. Note that since the matricies in (40) are recursive, finding the solution involves finding the fixed point of the system for a given Kalman gain (K) and pre-chosen upper limit (k^*) on the number of orders of expectations to include.

5 Proof of proposition 3.

For standard problems with imperfect common knowledge, where only the hierarchy of simple-average expectations is needed,[2] an arbitrarily accurate approximation of the full solution can be achieved by selecting a cut-off, k^*, and including all orders of expectation from zero to that cut-off, provided that

1. the importance attached to higher-*order* average expectations is decreasing in the order; and
2. the unconditional variance of higher-order average expectations are bounded from above.

The first of these is imposed by assumption. In the context of the model presented here, this amounts to a restriction on the coefficients in $\boldsymbol{\lambda}_1$.[3] The second is assured by the fact that agents are rational (Bayesian) and this is common knowledge. A proof of this is provided by Nimark (2011a), although it requires one minor extension here. Since I can write $X_t = E_t(j)[X_t] + X^{\text{err}}_{t|t}(j)$ and the variance of the two sides must be equal, I have

$$Var(X_t) = Var(E_t(j)[X_t]) + Var\left(X^{\text{err}}_{t|t}(j)\right)$$

where the covariance term on the right hand side can be ignored because j's rationality implies that her expectation must be orthogonal to her expectation error. This demonstrates that

$$Var(E_t(j)[X_t]) \leq Var(X_t)$$

The Kalman filter ensures that j's expectation must have a Moving Average representation incorporating linear combinations of the complete history of all shocks that enter her signals. For a simple average of this (lemma 2 in the nimark paper), any idiosyncratic shocks will necessarily sum to zero, ensuring that the simple-average expectation must have lower variance than that of any individual agent. For weighted averages of this, the idiosyncratic shocks will not sum to zero, but the variance of the weighted-average of those shocks will be less the variance of an individual shock as shown above in corollary 1 to proposition 1. It therefore must be that

$$Var\left(\overline{E}_t[X_t]\right) \leq Var\left(\overset{1:\sim}{E}_t[X_t]\right) \leq Var\left(\overset{2:\sim}{E}_t[X_t]\right) \leq \cdots \leq Var(E_t(j)[X_t]) \leq Var(X_t)$$

The recursive structure of X_t then establishes the result.

In addition, it is *also* necessary here to define a cut-off in the number of compound expectations to include (p^*). Analogously to the cut-off in higher orders of average expectation, the researcher's ability to deliver an arbitrarily accurate approximation requires that

1. the importance attached to higher-*weighted* expectations is decreasing in the weighting; and
2. the unconditional variance of higher-weighted average expectations are bounded from above.

The first of these is implied by the fact that each (next) higher weighted average expectation enters with a (further) lag and the underlying autoregressive process ensures that agents assign decreasing importance to older signals when considering their current expectation. The second was described above and is implied directly by corollary 1 to proposition 1.

[2] That is, where there is only one compound expectation of interest ($p = 1$).
[3] See section 4 of the main article for a typical example.

6 Implementation

Implementing a finite approximation with a cut-off, p^*, in the number of weighted-averages to include still requires that the programmer take a view on how to implement the the final weight. Recall from the main text that for the simplified model with no (lagged) public signal, the law of motion is

$$
\begin{aligned}
\boldsymbol{x}_t &= \rho\, \boldsymbol{x}_{t-1} && + \quad \boldsymbol{u}_t \\
\overline{E}_t\left[X_t\right] &= B\, \boldsymbol{x}_{t-1} + C\ \overline{E}_{t-1}\left[X_{t-1}\right] + D\ \overset{1:\sim}{E}_{t-1}\left[X_{t-1}\right] + H\boldsymbol{u}_t \\
\overset{1:\sim}{E}_t\left[X_t\right] &= B\, \boldsymbol{x}_{t-1} + C\ \overset{1:\sim}{E}_{t-1}\left[X_{t-1}\right] + D\ \overset{2:\sim}{E}_{t-1}\left[X_{t-1}\right] + H\boldsymbol{u}_t + Q\overset{1:\sim}{\boldsymbol{v}}_t \\
\overset{2:\sim}{E}_t\left[X_t\right] &= B\, \boldsymbol{x}_{t-1} + C\ \overset{2:\sim}{E}_{t-1}\left[X_{t-1}\right] + D\ \overset{3:\sim}{E}_{t-1}\left[X_{t-1}\right] + H\boldsymbol{u}_t + Q\overset{2:\sim}{\boldsymbol{v}}_t \\
&\ \vdots
\end{aligned}
$$

$$
\text{where} \quad
\begin{aligned}
B &= \boldsymbol{k}_p\, \rho & H &= \boldsymbol{k}_p \\
C &= F - BS_x - DT_{w_1} & Q &= q\, \boldsymbol{k}_p \\
D &= q\, \boldsymbol{k}_s\, \boldsymbol{\lambda}_1'
\end{aligned}
$$

with $\boldsymbol{k}_p$ being the Kalman gain applied to the private signal and $\boldsymbol{k}_s$ the Kalman gain applied to each social signal, so that the transition matrix for the full state therefore takes the following form:

$$
F = \begin{bmatrix}
\rho & 0 & 0 & 0 & \cdots \\
B & C & D & 0 & \\
B & 0 & C & D & \\
B & 0 & 0 & C & \ddots \\
\vdots & & & & \ddots
\end{bmatrix}
$$

For the p^{th}-weighted expectation, we have

$$
\begin{aligned}
\overset{p:\sim}{E}_t\left[X_t\right] &= BS_xX_{t-1} + CT_{w_p}X_{t-1} + DT_{w_{p+1}}X_{t-1} + shocks \\
&= \boldsymbol{k}_p\, \rho S_x X_{t-1} + \left(F - \boldsymbol{k}_p\, \rho S_x - q\, \boldsymbol{k}_s\, \boldsymbol{\lambda}_1' T_{w_1}\right) T_{w_p} X_{t-1} + q\, \boldsymbol{k}_s\, \boldsymbol{\lambda}_1' T_{w_{p+1}} X_{t-1} + shocks \\
&= \left(\boldsymbol{k}_p\, \rho S_x + \left(F - \boldsymbol{k}_p\, \rho S_x\right) T_{w_p}\right) X_{t-1} + q\, \boldsymbol{k}_s\, \boldsymbol{\lambda}_1' \left(T_{w_{p+1}} - T_{w_1} T_{w_p}\right) X_{t-1} + shocks
\end{aligned}
$$

When considering the expectations of agents p levels deep in the network, the component derived from consideration of agents $p+1$ levels deep is captured in the term $q\, \boldsymbol{k}_s\, \boldsymbol{\lambda}_1' \left(T_{w_{p+1}} - T_{w_1} T_{w_p}\right) X_{t-1}$. For the final weighting in the simulation, two clear possibilities are apparent:

- For the final weight, use $q\, \boldsymbol{k}_s\, \boldsymbol{\lambda}_1' \left(T_{w_{p^*}} - T_{w_1} T_{w_{p^*}}\right) X_{t-1}$
- For all weights $\Psi q\, \boldsymbol{k}_s\, \boldsymbol{\lambda}_1' \left(T_{w_{p+1}} - T_{w_1} T_{w_p}\right) X_{t-1}$ and have $\Psi = 1$ for $p < p^*$ and $\Psi = 0$ for $p = p^*$

The first option implies that agents treat competitors p and $p+1$ levels deep in the network the same, and know that all other agents take the same approach. The second option implies that agents suppose that competitors p levels deep in the network do not observe anybody so their information comes only from their public/private signals. Both options must be equivalent as $p^* \to \infty$ and, in practice, are seen to produce highly similar results.

The attached Matlab code provides an implementation of the model that uses a third alternative:

- For the final weight, use $\Psi q\, \boldsymbol{k}_s\, \boldsymbol{\lambda}_1' \left(T_{w_{p^*}} - T_{w_1} T_{w_{p^*}} \right) X_{t-1}$ where $\Psi = 1 + \epsilon$.

which assumes that agents treat competitors p and $p+1$ levels deep in the network the same *and* artificially forces them to place slightly more weight on them when constructing the Kalman filter in order to crudely capture the unsimulated higher-weighted expectations. Doing this improves the implementation's robustness to numerical instability and allows simulations with higher numbers of observees (q).

Numerical instability

Although equations (35) and (40) provide the algorithm through which to iterate, as written they are extremely memory intensive and prone to numerical instability. This problem worsens as q increases and, for moderate-to-high persistence in the underlying state, the solution can only be found for very low values of q.

Without recourse to standard UD-factorisation techniques (see the main text), then in addition to the avoidance of stacking the state vector already implemented and the implementation of p^* mentioned above, I also deploy the following techniques to improve the algorithm's performance:

Avoid unnecessary iteration

As mentioned above, the network learning problem involves finding convergent solutions to the filter and the law of motion, each taking the other as given. In principle, the fixed point may therefore be found by finding the convergent result of one within each iteration of the other – for example:

```
repeat
    Update the filter by one step using equation (35)
    repeat
        Update the law of motion by one step using equation (40)
    until the law of motion converges
until the filter converges
```

This set-up is $O\left(n^2\right)$, however, even before examining the complexity of the one-step processes, and in practice is more likely to suffer from numerical stability issues. Instead, for a given set of signals, I find the fixed point by updating the filter and the law of motion incrementally within the same loop:

```
repeat
    Update the filter by one step using equation (35)
    Update the law of motion by one step using equation (40)
until both the filter and the law of motion converge
```

Avoid temporary creation of unnecessarily large matrices

The solution as presented above (see equations 35a and 35b) involves the temporary creation (and multiplication) of matrices that are $(2+q) \times N$ square, where N is the size of X_t and q is the number of other agents observed.

The implementation presented in the attached Matlab code keeps the public/private signals and the social signals separate (i.e. it breaks the M_* and N_* matrices into their constituent components) to avoid this and to exploit the fact that each social signal will be treated identically.

Pay close attention to operation order

Because matrix addition and subtraction are of order $O\left(n^2\right)$ while (naive) matrix multiplication and inversion are of order $O\left(n^3\right)$, the order in which expressions are calculated can affect the number of operations required.

For example, although mathematically equivalent, the computational complexity of calculating $(A+B)\times C$ is *less* than that of $(A\times C)+(B\times C)$ because the former involves only a single multiplication.

7 Extending the model to dynamic actions

We here consider an illustrative example of extending the model of this chapter to consideration of dynamic actions. In particular, we allow agents' decision rules to be slightly more general, with an inclusion of agents' expectations regarding the next-period average action. That is, we suppose that individual decisions are made according to the following rule:

$$g_t(i) = \boldsymbol{\alpha}' \boldsymbol{s}_t^p(i) + \boldsymbol{\eta}_x' E_t(i)\left[X_t\right] + \eta_y E_t(i)\left[\overline{g}_t\right] + \eta_z E_t(i)\left[\overline{g}_{t+1}\right] \tag{41}$$

where agents' private signals are formed as

$$\boldsymbol{s}_t^p(i) = B\boldsymbol{x}_t + Q\boldsymbol{v}_t(i)$$

We retain the assumption that the underlying state follows an AR(1) process:

$$\boldsymbol{x}_t = A\boldsymbol{x}_{t-1} + P\boldsymbol{u}_t$$

and still suppose that the full hierarchy of expectations regarding the underlying state is given by:

$$X_t = \mathbb{E}_t^{(0:\infty)}\left[\boldsymbol{x}_t\right]$$

Our goal is to show that $g_t(i)$ may be expressed in the general form

$$g_t(i) = \boldsymbol{\lambda}_0' w_{t-1} + \boldsymbol{\lambda}_2' X_t + \boldsymbol{\lambda}_1' E_t(i)\left[X_t\right] + \boldsymbol{\lambda}_3' \boldsymbol{v}_t(i)$$

To do this, we start by taking the simple average of equation (41) to give:

$$\overline{g}_t = \boldsymbol{\alpha}' B\boldsymbol{x}_t + \boldsymbol{\eta}_x' \overline{E}_t\left[X_t\right] + \eta_y \overline{E}_t\left[\overline{g}_t\right] + \eta_z \overline{E}_t\left[\overline{g}_{t+1}\right]$$

To keep the notation clean, define $\theta_t \equiv \boldsymbol{\alpha}' B\boldsymbol{x}_t + \boldsymbol{\eta}_x' \overline{E}_t\left[X_t\right]$ so that

$$\overline{g}_t = \theta_t + \eta_y \overline{E}_t\left[\overline{g}_t\right] + \eta_z \overline{E}_t\left[\overline{g}_{t+1}\right]$$

We now substitute this equation back into itself in the second element ($\eta_y \overline{E}_t\left[\overline{g}_t\right]$):

$$\overline{g}_t = \theta_t + \eta_y \overline{E}_t\left[\theta_t\right] + \eta_y^2 \overline{E}_t^{(2)}\left[\overline{g}_t\right] + \eta_z \overline{E}_t\left[\overline{g}_{t+1}\right] + \eta_y \eta_z \overline{E}_t^{(2)}\left[\overline{g}_{t+1}\right]$$

Repeating this process, in the limit (and using the fact that $\eta_y \in (0,1)$ and assuming that average expectations don't explode), this becomes:

$$\overline{g}_t = \left(\sum_{k=0}^{\infty} \eta_y^k \overline{E}_t^{(k)}\left[\theta_t\right]\right) + \left(\eta_z \sum_{k=1}^{\infty} \eta_y^{k-1} \overline{E}_t^{(k)}\left[\overline{g}_{t+1}\right]\right)$$

now briefly consider θ_t and simple-average expectations of θ_t. We can write that:

$$\begin{aligned}\theta_t &= \boldsymbol{\alpha}' B\boldsymbol{x}_t + \boldsymbol{\eta}_x' \overline{E}_t^{(1)}[X_t] \\ \overline{E}_t^{(1)}[\theta_t] &= \boldsymbol{\alpha}' B\overline{E}_t^{(1)}[\boldsymbol{x}_t] + \boldsymbol{\eta}_x' \overline{E}_t^{(2)}[X_t] \\ \overline{E}_t^{(2)}[\theta_t] &= \boldsymbol{\alpha}' B\overline{E}_t^{(2)}[\boldsymbol{x}_t] + \boldsymbol{\eta}_x' \overline{E}_t^{(3)}[X_t] \\ &\dots\end{aligned}$$

next, suppose that the matrix T_s selects the simple-average expectation of X_t from X_t:

$$\overline{E}_t^{(1)}[X_t] = T_s X_t$$

and that the matrix S selects $\boldsymbol{x}_t$ from X_t (obviously $S = \begin{bmatrix} I_l & 0_{l\times\infty}\end{bmatrix}$ where l is the number of elements in $\boldsymbol{x}_t$):

$$\boldsymbol{x}_t = SX_t$$

Then we can write:

$$\begin{aligned}\theta_t &= (\boldsymbol{\alpha}' BS + \boldsymbol{\eta}_x' T_s) X_t \\ \overline{E}_t^{(1)}[\theta_t] &= (\boldsymbol{\alpha}' BS + \boldsymbol{\eta}_x' T_s) T_s X_t \\ \overline{E}_t^{(2)}[\theta_t] &= (\boldsymbol{\alpha}' BS + \boldsymbol{\eta}_x' T_s) T_s^2 X_t \\ &\dots\end{aligned}$$

or, in general,

$$\overline{E}_t^{(k)}[\theta_t] = (\boldsymbol{\alpha}' BS + \boldsymbol{\eta}_x' T_s) T_s^k X_t$$

The average period-t action can therefore be written as

$$\begin{aligned}\overline{g}_t &= (\boldsymbol{\alpha}' BS + \boldsymbol{\eta}_x' T_s)\left(\sum_{k=0}^{\infty}(\eta_y T_s)^k\right) X_t + \eta_z \sum_{k=1}^{\infty}\eta_y^{k-1}\overline{E}_t^{(k)}[\overline{g}_{t+1}] \\ &= (\boldsymbol{\alpha}' BS + \boldsymbol{\eta}_x' T_s)(I - \eta_y T_s)^{-1} X_t + \eta_z \sum_{k=1}^{\infty}\eta_y^{k-1}\overline{E}_t^{(k)}[\overline{g}_{t+1}] \\ &= \boldsymbol{\beta}' X_t + \eta_z \sum_{k=1}^{\infty}\eta_y^{k-1}\overline{E}_t^{(k)}[\overline{g}_{t+1}]\end{aligned}$$

where $\boldsymbol{\beta}' \equiv (\boldsymbol{\alpha}' BS + \boldsymbol{\eta}_x' T_s)(I - \eta_y T_s)^{-1}$. Next, substitute this back into itself for the next-period average action:

$$\begin{aligned}\overline{g}_t &= \boldsymbol{\beta}' X_t + \eta_z \sum_{k=1}^{\infty}\eta_y^{k-1}\overline{E}_t^{(k)}\left[\boldsymbol{\beta}' X_{t+1} + \eta_z \sum_{l=1}^{\infty}\eta_y^{l-1}\overline{E}_{t+1}^{(l)}[\overline{g}_{t+2}]\right] \\ &= \boldsymbol{\beta}' X_t + \eta_z \sum_{k=1}^{\infty}\eta_y^{k-1}\boldsymbol{\beta}'\overline{E}_t^{(k)}[X_{t+1}] + \eta_z \sum_{k=1}^{\infty}\eta_y^{k-1}\overline{E}_t^{(k)}\left[\eta_z \sum_{l=1}^{\infty}\eta_y^{l-1}\overline{E}_{t+1}^{(l)}[\overline{g}_{t+2}]\right]\end{aligned}$$

next, we use the following conjectured aspect of the law of motion for X_t:

$$E_t(i)[X_{t+1}] = E_t(i)[FX_t]$$

for some matrix of parameters F. This implies that

$$\overline{E}_t^{(k)}[X_{t+1}] = F\overline{E}_t^{(k)}[X_t]$$

and hence that

$$\begin{aligned}\overline{g}_t &= \boldsymbol{\beta}' X_t + \eta_z \boldsymbol{\beta}' F \sum_{k=1}^{\infty} \eta_y^{k-1} \overline{E}_t^{(k)} [X_t] + \eta_z \sum_{k=1}^{\infty} \eta_y^{k-1} \overline{E}_t^{(k)} \left[\eta_z \sum_{l=1}^{\infty} \eta_y^{l-1} \overline{E}_{t+1}^{(l)} [\overline{g}_{t+2}] \right] \\ &= \boldsymbol{\beta}' X_t + \eta_z \boldsymbol{\beta}' F \left(\sum_{k=1}^{\infty} \eta_y^{k-1} T_s^k \right) X_t + \eta_z \sum_{k=1}^{\infty} \eta_y^{k-1} \overline{E}_t^{(k)} \left[\eta_z \sum_{l=1}^{\infty} \eta_y^{l-1} \overline{E}_{t+1}^{(l)} [\overline{g}_{t+2}] \right] \\ &= \boldsymbol{\beta}' X_t + \eta_z \boldsymbol{\beta}' F T_s (I - \eta_y T_s)^{-1} X_t + \eta_z \sum_{k=1}^{\infty} \eta_y^{k-1} \overline{E}_t^{(k)} \left[\eta_z \sum_{l=1}^{\infty} \eta_y^{l-1} \overline{E}_{t+1}^{(l)} [\overline{g}_{t+2}] \right]\end{aligned}$$

next, expand the $\overline{g}_{t+2}$ term to give

$$\begin{aligned}\overline{g}_t &= \boldsymbol{\beta}' X_t + \eta_z \boldsymbol{\beta}' F T_s (I - \eta_y T_s)^{-1} X_t \\ &\quad + \eta_z \sum_{k=1}^{\infty} \eta_y^{k-1} \overline{E}_t^{(k)} \left[\eta_z \sum_{l=1}^{\infty} \eta_y^{l-1} \overline{E}_{t+1}^{(l)} \left[\boldsymbol{\beta}' X_{t+2} + \eta_z \sum_{m=1}^{\infty} \eta_y^{m-1} \overline{E}_{t+2}^{(m)} [\overline{g}_{t+3}] \right] \right] \\ &= \boldsymbol{\beta}' X_t \\ &\quad + \eta_z \boldsymbol{\beta}' F T_s (I - \eta_y T_s)^{-1} X_t \\ &\quad + \boldsymbol{\beta}' \left(\eta_z F T_s (I - \eta_y T_s)^{-1} \right)^2 X_t \\ &\quad + \eta_z \sum_{k=1}^{\infty} \eta_y^{k-1} \overline{E}_t^{(k)} \left[\eta_z \sum_{l=1}^{\infty} \eta_y^{l-1} \overline{E}_{t+1}^{(l)} \left[\eta_z \sum_{m=1}^{\infty} \eta_y^{m-1} \overline{E}_{t+2}^{(m)} [\overline{g}_{t+3}] \right] \right]\end{aligned}$$

Continued substitution then arrives at:

$$\overline{g}_t = \boldsymbol{\beta}' \sum_{j=0}^{\infty} \left(\eta_z F T_s (I - \eta_y T_s)^{-1} \right)^j X_t$$

which, in turn, becomes

$$\overline{g}_t = \underbrace{(\boldsymbol{\alpha}' B S + \boldsymbol{\eta}_x' T_s) (I - \eta_y T_s)^{-1} \left(I - \eta_z F T_s (I - \eta_y T_s)^{-1} \right)^{-1}}_{\equiv \boldsymbol{a}'} X_t$$

Using this simple expression of $\overline{g}_t = \boldsymbol{a}' X_t$, we can substitute it back into the agents' individual decision rule to obtain

$$\begin{aligned}g_t(i) &= \boldsymbol{\alpha}' (B \boldsymbol{x}_t + Q \boldsymbol{v}_t(i)) + (\boldsymbol{\eta}_x' + \eta_y \boldsymbol{a}' + \eta_z \boldsymbol{a}' F) E_t(i) [X_t] \\ &= \underbrace{\boldsymbol{\alpha}' B}_{\boldsymbol{\lambda}_2'} \boldsymbol{x}_t + \underbrace{(\boldsymbol{\eta}_x' + \eta_y \boldsymbol{a}' + \eta_z \boldsymbol{a}' F)}_{\gamma_3'} E_t(i) [X_t] + \underbrace{\boldsymbol{\alpha}' Q}_{\gamma_4'} \boldsymbol{v}_t(i)\end{aligned}$$

which is now in the necessary form. As an aside, taking a simple average of this gives

$$\overline{g}_t = \boldsymbol{\alpha}' B S X_t + (\boldsymbol{\eta}_x' + \eta_y \boldsymbol{a}' + \eta_z \boldsymbol{a}' F) \overline{E}_t [X_t]$$

which implies the following constraint on the coefficients of the decision rule ($\boldsymbol{\alpha}$, $\boldsymbol{\eta}_x$, $\boldsymbol{\eta}_y$, $\boldsymbol{\eta}_z$) and the expectation transition matrix (F):

$$\boldsymbol{a}' = \boldsymbol{\alpha}' B S + (\boldsymbol{\eta}_x' + \eta_y \boldsymbol{a}' + \eta_z \boldsymbol{a}' F) T_s$$

References

Nimark, K. (2011a): "Dynamic Higher Order Expectations," Universitat Pompeu Fabra Economics Working Papers No 1118.

——— (2011b): "A low dimensional Kalman Filter for systems with lagged observables," mimeo.