Incomplete information and the idiosyncratic foundations of aggregate volatility

John Barrdear

27th July 2013

A thesis submitted to the Department of Economics of the London School of Economics and Political Science for the degree of Doctor of Philosophy

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work.

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that this thesis consists of 69,734 words.[1]

[1] Calculated by TexMaker, v3.5.

Abstract

This thesis considers two interrelated themes: the emergence of aggregate volatility from idiosyncratic shocks and optimisation under incomplete information when, for reasons of strategic complementarity, agents are interested in both simple and weighted averages of their competitors' actions.

I first develop a model of Bayesian social learning over a network. Unlike earlier literature that abandons one of the assumptions that agents (a) act repeatedly; (b) are rational; and (c) face strategic complementarities, I obtain tractability for arbitrarily large networks by also assuming that agents do not know the full structure of the network, but do know its link distribution. An AR(1) process for the underlying state induces an ARMA(1,1) process for the hierarchy of expectations, with current and lagged weighted averages of agents' idiosyncratic shocks entering at an aggregate level. For sufficiently irregular networks, these shocks do not wash out, thus causing persistent aggregate effects.

I next apply this to firms' price-setting problem, demonstrating that even when firms possess complete price flexibility, network learning induces considerable persistence in aggregate variables following monetary and real shocks and that network shocks plausibly represent a source of aggregate economic volatility.

Finally, I explore price setting under monopolistic competition when facing Trans-Log preferences. I solve explicitly for a firm's best-response pricing rule under full information and show that in partial equilibrium under incomplete information, larger firms will focus on their marginal costs while smaller firms will place more weight on changes in consumer preferences and competitors' prices. In general equilibrium, I estimate the effect of two distinct sources of real rigidity that emerge from Trans-Log preferences: the well-known curvature in demand and the dramatic increase in complexity of firms' signal-extraction problems. With non-uniform preferences, the model represents another channel through which idiosyncratic shocks can cause aggregate volatility.

Acknowledgements

This thesis would not have been possible without the extensive support and assistance of a great many people. I would first like to thank my supervisor, Kevin Sheedy, for his incredible willingness to spend hours at a time patiently trying to understand what I was saying and his encouragement in my continuing to pursue my goal even when things looked bleakest. I am also thankful to the rest of the LSE faculty, particularly Andrea Prat, Silvana Tenreyro, Francesco Caselli, Francesco Nava and Wouter DenHaan for their encouragement and invaluable feedback during seminars.

Of my fellow PhD candidates, James Hansen, Gianni LaCava, Victor Ortego-Marti, Matthew Skellern, Dimitri Szerman and Luke Miner were exceptional in their efforts as sounding boards and their companionship as friends. I thank them for the laughs, the support, the willingness to listen to me rant and the odd game of poker over the years. Luke and Dimitri, especially, were my brothers in arms in my time at LSE and I could not have completed this work without them.

My family have been my bedrock in tackling this PhD. I am, and always will be, grateful to my parents for the most magnificent upbringing a son could hope for and the instillation of the belief that all things are possible; to James for his sense of humour, his sense of perspective and his ever-ready ear; and to Alec for his encouragement and the endless debates, hammering a sense of ethics into my otherwise dry economics. To all of them, I owe thanks for the various admonishments, prods, pushes and occasional kicks up the bum to get this over the line.

Finally, and most importantly, I must extend my warmest thanks to my wife, Daniela. It is impossible for me to express, let alone repay, my debt of gratitude to Dani for the love, encouragement and patient support she has shown me throughout my time at LSE. Her tolerance of my endless weekends spent working, her willingness to listen to my obsessive rambling about the most esoteric minutia and the sheer strength of her belief in me have proven, time and again, to be without bound.

Angel, this is all for you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and common definitions

1.1 Introduction

This thesis is interested in two fundamental questions of macroeconomics. First, what are the underlying causes of observed volatility in aggregate variables? From where do the shocks arise and of what are they comprised? Second, what explains the magnitude and persistence of the effects of aggregate shocks on the macroeconomy? Is it possible for those effects to persist beyond the shock itself, or beyond the time when agents in the economy successfully identify it?

These are by no means new questions – macroeconomists have been grappling with them for decades – but they remain open questions of active research and recent work has brought each of them (and previous attempts at answering them) into a new light.

The first question may to some extent strike readers as odd, as it is natural to suppose that aggregate volatility must be caused by aggregate shocks, at least when considering a linear model. However, recent work by Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Saleh (2012) and Gabaix (2011) has raised the possibility of idiosyncratic shocks having aggregate effects when their economic relevance emerges through distributions that are *fat tailed* (i.e. when certain agents' shocks or actions play a disproportionately important role in aggregation, while the majority of agents receive a disproportionately low weight). A central contribution of this thesis is to add to this burgeoning literature by outlining and characterising two new settings in which this result applies. In both cases, the result relies on the presence of a distribution that is *asymptotically non-uniform*, in that it remains sufficiently far from uniform even as the number of agents over which it applies grows arbitrarily large. Section 1.3 below offers a formal definition of such a distribution, demonstrates why it allows for idiosyncratic shocks to have aggregate effects and discusses the class of distributions that meet such a characteristic.

The second question is of particular importance to a policy maker seeking to dampen aggregate volatility in an economy. Statistical analyses, relying only minimally, if at all, on economic theory, have suggested that the effects on real GDP and aggregate prices from identified aggregate shocks, both real and nominal, are significant and quite persistent. Understanding why this is so, particularly for nominal shocks, is therefore of critical importance.

The most obvious way to model rigidity in aggregate prices is naturally to suppose that individual prices are themselves in some way "sticky." This approach was supported by early quantitative evidence, which suggested that prices, once set, generally remained fixed for 12 months (Taylor, 1999). With this in mind, the canonical New Keynesian model was developed by assuming that firms follow simple *time-based* pricing rules. That is, that firms choose only the magnitude of any price changes, while the timing of those changes is determined exogenously. Standard approaches here are the staggered pricing rule of Taylor (1980), in which firms' prices remain fixed for n periods and a fraction $1/n$ of firms update their prices each period, and the model of Calvo (1983), in which a random fraction of firms are able to update in each period. Since simple models of *state-based* pricing, such as those proposed by Rotemberg (1982), imply similar Phillips curves under linearisation, it was generally held that the analytically simple time-based rules were sufficient.

However, a variety of challenges to simple time-based pricing rules have arisen in recent years. From a theoretical perspective, models that exogenously impose the timing of price changes would appear to fail the Lucas Critique by assuming, rather than deriving, policy invariance in agents' actions (e.g. Plosser, 2012). Indeed, modern models of state-based pricing (i.e. that look both at the magnitude of price changes and whether and when to change) have consistently shown that *selection effects* work against any persistence of aggregate effects on real variables, with the firms that elect to adjust their prices following a monetary policy shock being precisely those that make the largest adjustment (e.g. Gertler and Leahy, 2008).

Empirically, a variety of modern studies of microeconomic price changes have demonstrated that a great many prices are remarkably short lived. The seminal work of Bils and Klenow (2004), for example, found that the median duration of prices in CPI data in the United States was 4.3 months, an update frequency almost three times higher than previously thought. More recent work by Klenow and Kryvtsov (2008) found a median duration of only 3.7 months and highlighted that a significant fraction of goods' prices are changed on a weekly basis.

With prices apparently quite flexible and selection effects ensuring minimal real effects via what individual price stickiness remains, it is therefore conceptually necessary to develop models of real rigidity – a "contract multiplier," in the words of Taylor (1980) – to explain the sluggish responses observed in aggregate price indicies following monetary shocks.

One key approach to achieving real rigidities in firms' pricing is to suppose that they have imperfect access to information, an idea that dates to Lucas (1972) and Phelps (1984). Modern models in this field fall broadly into three categories:

1. **Sticky Information.** First, as argued by Mankiw and Reis (2002), firms may have access to information only infrequently, so that in aggregate, they only respond to a policy change gradually. Reis (2006) provides a microfoundation for this idea, arguing that information processing costs (as distinct from classic menu costs) make it optimal for firms to delay their updates.
2. **Rational Inattention.** Second, as suggested by Sims (2003), firms may be subject to an information processing constraint, whereby there is a limit to the amount of information they can accommodate, irrespective of the cost of doing so. In such a case, it is rational for firms to access information in every period, but to select which signals to observe.
3. **Incomplete Information.** Finally, as proposed by Woodford (2003), even if firms are free to update their information sets every period and face no constraints in doing so, they may be subject to structural imperfections in their signals. When noisy signals of hidden aggregate states are combined with strategic complementarity, this gives rise to what Woodford termed "Imperfect

Common Knowledge," with firms needing to consider the higher-order beliefs of their competitors.

The work of this thesis fits squarely in the latter of these categories, extending the problem to scenarios where firms must consider not just the simple average of other firms' expectations, but weighted averages as well. The sizes of firms' state vectors become much larger in these settings and their signal extraction problems become correspondingly more difficult, thus leading to more sluggish responses in prices following aggregate shocks, even when firms are free to update their information sets and their prices every period.

Section 1.2 below offers some necessary definitions relating to higher-order expectations when there are multiple expectations of interest and illustrates the explosion in the size of the state vector that they can imply.

1.2 Higher-order expectations

The near-ubiquitous treatment of higher-order expectations in the macroeconomic literature to date[1] has been to consider only the hierarchy of *simple average* expectations. That is, to consider settings in which 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 $\overline{E}_t\left[\cdot\right] \equiv \int_0^1 E_t\left(i\right)\left[\cdot\right] di$.

It is important to realise that this is a modelling choice only, made for analytical convenience. It is easy to envisage scenarios where other, more complex hierarchies of beliefs are relevant and this thesis occupies itself with two such examples. The first, dealt with in chapters 2 and 3, involves rational learning over a network, in which 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. The second, addressed in chapter 4, examines firms' price-setting problem when household preferences are non-uniform, so that every firm must consider two separate aggregations of belief, one of which is firm-specific.

To model these fully, we therefore first provide a generalised definition of a hierarchy of expectations.

Definition 1. *A **compound expectation** is a weighted sum of all agents' expectations. 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 $\mathcal{E}_t\left[\boldsymbol{x}_t\right] \equiv \begin{bmatrix} E\left[\boldsymbol{x}_t | \mathcal{I}_t\left(1\right)\right] & \cdots & E\left[\boldsymbol{x}_t | \mathcal{I}_t\left(N\right)\right]\end{bmatrix}$ be the $(m \times N)$ matrix containing all agents' expectations of the same. Let $\boldsymbol{w}$ be an $(N \times 1)$ vector of weights across all agents 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:*

$$E_{\boldsymbol{w},t}\left[\boldsymbol{x}_t\right] \equiv \mathcal{E}_t\left[\boldsymbol{x}_t\right] \boldsymbol{w} \tag{1.1}$$

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. $\boldsymbol{w}_B = \begin{bmatrix} \mathbf{0}' & 1 & \mathbf{0}'\end{bmatrix}'$).

Definition 2. *Let $W \equiv \begin{bmatrix} \boldsymbol{w}_A & \boldsymbol{w}_B & \cdots \end{bmatrix}$ be the $(N \times p)$ matrix formed of all weights of interest in a given problem and p be the number of those weights (i.e. the number*

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

*of columns in W). We then define **higher-order expectations** as follows, using a blackboard-bold $\mathbb{E}^{(k)}$ to denote the vector containing all expectations of the k-th order:*

$$\mathbb{E}_t^{(0)}[\boldsymbol{x}_t] \equiv \boldsymbol{x}_t$$
$$\mathbb{E}_t^{(k)}[\boldsymbol{x}_t] \equiv \begin{bmatrix} E_{\boldsymbol{w}_A,t}\left[\mathbb{E}_t^{(k-1)}[\boldsymbol{x}_t]\right] \\ E_{\boldsymbol{w}_B,t}\left[\mathbb{E}_t^{(k-1)}[\boldsymbol{x}_t]\right] \\ \vdots \end{bmatrix} = vec\left(\mathcal{E}_t\left[\mathbb{E}_t^{(k-1)}[\boldsymbol{x}_t]\right]W\right) \ \forall k \geq 1 \tag{1.2}$$

Note that if we are interested in p different compound expectations, there are p^k different permutations of k-th order expectations. For example, if $\boldsymbol{x}_t$ is scalar and $p = 2$, then the vector describing the set of second-order expectations will be of size (4×1) and arranged in the following way:

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

Definition 3. *A **hierarchy of expectations**, from order 0 to k, is defined recursively as:*

$$\mathbb{E}_t^{(0:k)}[\boldsymbol{x}_t] = \begin{bmatrix} \boldsymbol{x}_t \\ E_{\boldsymbol{w}_A,t}\left[\mathbb{E}_t^{(0:k-1)}[\boldsymbol{x}_t]\right] \\ E_{\boldsymbol{w}_B,t}\left[\mathbb{E}_t^{(0:k-1)}[\boldsymbol{x}_t]\right] \\ \vdots \end{bmatrix} \tag{1.3}$$

Note that 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 $p = 2$, the hierarchies $(0:1)$ and $(0:2)$ are given by:

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

The benefit of depicting hierarchies in this recursive manner is that it becomes simple to extract sub-hierarchies comprised of a single compound expectation. For example, if $\boldsymbol{w}_A = \begin{bmatrix} \frac{1}{N} & \cdots & \frac{1}{N} \end{bmatrix}'$ so that $E_{\boldsymbol{w}_A,t}\left[\boldsymbol{x}_t\right] = \overline{E}_t\left[\boldsymbol{x}_t\right]$ is the average expectation, the sub-hierarchy of $\overline{\boldsymbol{x}}_t^{(0:k)} \equiv \left[\boldsymbol{x}_t', \overline{E}_t\left[\boldsymbol{x}_t'\right], \overline{E}_t\left[\overline{E}_t\left[\boldsymbol{x}_t'\right]\right], \cdots\right]'$ may be extracted as:

$$\overline{\boldsymbol{x}}_t^{(0:k)} = \begin{bmatrix} I & 0 \end{bmatrix} \mathbb{E}_t^{(0:k)}\left[\boldsymbol{x}_t\right]$$

In all of the models in this thesis, the expectation hierarchy $\mathbb{E}_t^{(0:\infty)}\left[\boldsymbol{x}_t\right]$ will represent the unknown state vector about which agents attempt to learn.

1.2.1 Size of the expectation hierarchy

Although not of particular importance in theory, the size of the state vector of interest is of crucial importance if any model is to be simulated. It is clear that if $\boldsymbol{x}_t$ contains m elements, then $\mathbb{E}_t^{(k)}\left[\boldsymbol{x}_t\right]$ – the set of k-th order expectations – will contain mp^k distinct elements. However, it is worth emphasising that it does not in general follow that the hierarchy $\mathbb{E}_t^{(0:k^*)}\left[\boldsymbol{x}_t\right]$ will contain $m\left(\sum_{k=0}^{k^*} p^k\right)$ unique elements. This is because if one of the compound expectations, say $E_{\boldsymbol{w}_B}\left[\cdot\right]$, is formed from a single information set – i.e. a single agent's expectation – then the law of iterated expectations implies that $E_{\boldsymbol{w}_B,t}\left[E_{\boldsymbol{w}_B,t}\left[\boldsymbol{x}_t\right]\right] = E_{\boldsymbol{w}_B,t}\left[\boldsymbol{x}_t\right]$.

In general, when $q\left(\leq p\right)$ is the number of individual expectations in W, the number of unique elements in the hierarchy $\mathbb{E}_t^{(0:k^*)}[\boldsymbol{x}_t]$ will be given by:[2]

$$N(m,p,q,k^*) = m\left(p^{k^*} + \sum_{k=0}^{k^*-1}\left(p^k - q\sum_{s=0}^{k} p^s\right)\right) \tag{1.4}$$

with $N(m,p,0,k^*) = m\left(\sum_{k=0}^{k^*} p^k\right)$. Nevertheless, even when $q = p$, it should be readily apparent that size of an expectation hierarchy explodes in both p (the number of compound expectations) and k^* (the highest order in expectations). Figure 1.1 illustrates this point, plotting the size of the hierarchy when $q = 0$ and $m = 1$.

A state vector of infinite dimension need not be a problem, *per se*, provided that the researcher is able to make a reasonable approximation of agents' actions by restricting attention to a finite subset of the state. In most models – including those of this thesis – imposing a finite upper limit, k^*, on the number of orders of expectation will be acceptable as in order to ensure stability in agent actions, decreasing weight is placed on higher order expectations.

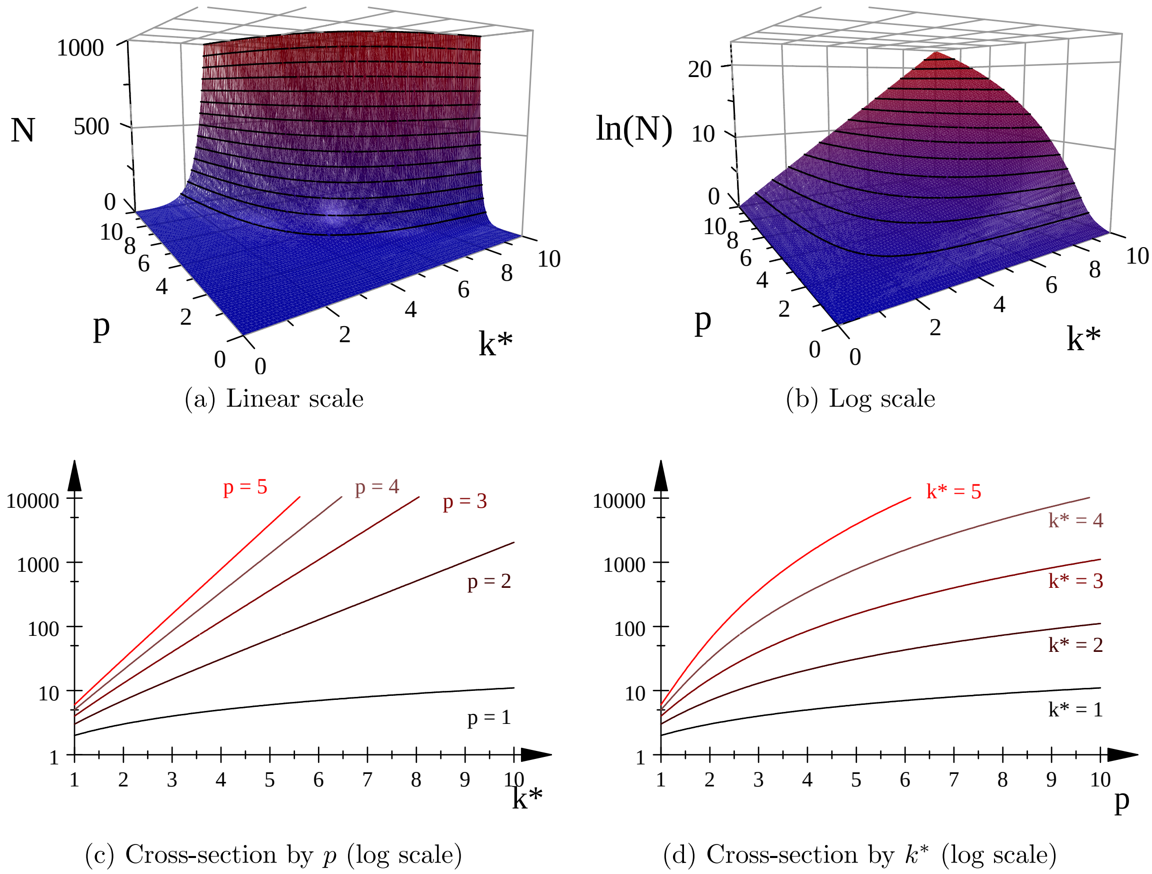
On the other hand, allowing the number of relevant compound expectations to increase can be more problematic as there is rarely, if ever, an obvious reason for weighting them differently. Existing work in the macroeconomic literature has generally avoided this difficulty by limiting attention to problems that implicitly assume that $p = 1$ (in particular, that all agents care only about the simple average expectation of their competitors).

This avenue is not available when considering learning via networks, however, where it is typically the case that p is given by the number of agents in the network, or in the case of asymmetric preferences examined in the final chapter, where p is given by the number of firms in the joint demand system.

[2]

$$m\left(\underbrace{[1]}_{\text{0-th order}} + \underbrace{[p]}_{\text{1-st order}} + \underbrace{[p^2-q]}_{\text{2-nd order}} + \underbrace{[p*(p^2-q)-q]}_{\text{3-rd order}} + \underbrace{[p*(p*(p^2-q)-q)-q]}_{\text{4-th order}} + \cdots\right)$$

$$= m\left(\left(\sum_{k=0}^{k^*} p^k\right) - q\left(\sum_{k=0}^{k^*-1}\sum_{s=0}^{k} p^s\right)\right)\text{, which rearranges to the equation in the text}$$



(a) Linear scale

(b) Log scale

(c) Cross-section by p (log scale)

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

Figure 1.1: The number of elements in an expectation hierarchy ($q = 0$, $m = 1$)

1.3 Asymptotically non-uniform distributions

The second key theme of this thesis is an illustration of how idiosyncratic shocks need not "wash out" and may, instead, induce aggregate volatility in an economic context. Fundamentally, this implies an exploration of settings in which the standard law of large numbers does not apply which, in turn, implies that the models must consider *weighted* sums of agents' idiosyncratic shocks.

Identifying laws of large numbers for weighted sums of i.i.d. random variables (i.e. the limiting behaviour of $\sum_{i=1}^{N} a_{N,i} X_i$ when $E[X] = 0$) remains an area of active research.[3] However, we do not require an exact characterisation of the necessary conditions for a weighted sum to converge to zero, as there is a broad range of functions for the weights under which a weighted sum will *not* converge to zero. In particular, it is sufficient to suppose that the weights are asymptotically non-uniform:

Definition 4. *Let Φ_N be a discrete distribution with corresponding p.d.f.*[4] $\phi_N(i)$. *Let* $\zeta(N) \equiv \sum_{i=1}^{N} \phi_N(i)^2$ *be the Herfindahl-Hirschman index for the same. The distribution 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 appreciate how such a distribution is sufficient to ensure that idiosyncratic shocks do not wash out, suppose that each agent receives an independent, mean zero shock drawn from a common Gaussian distribution (i.e. one fully characterised by its first and second moments):

$$\boldsymbol{v}(i) \sim N(\mathbf{0}, \Sigma_{vv}) \;\; \forall i$$

[3] See, for example, Wu (1999), Sung (2001) or Cai (2006).

[4] Strictly, for a discrete distribution, it is a probability *mass* function. But since we will concern ourselves only with the limiting case of $N \to \infty$ and assume that they are indexed uniformly from zero to one so that the distribution becomes continuous, we stick with the conventional nomenclature.

and consider the setting where it is not the simple average of agents' shocks that matters, but a *weighted* average:

$$\widetilde{\boldsymbol{v}}_N \equiv \sum_{i=1}^{N} \boldsymbol{v}(i)\,\phi_N(i) \text{ where } \phi_N(i) \in (0,1) \text{ and } \sum_{i=1}^{N} \phi_N(i) = 1$$

Since $\widetilde{\boldsymbol{v}}_N$ is a linear combination of mean-zero Gaussian variables, it must itself have a Normal distribution with a mean of zero. Its variance will then be given by:

$$\begin{aligned} Var\left[\widetilde{\boldsymbol{v}}_N\right] &= Var\left[\sum_{i=1}^{N} \boldsymbol{v}(i)\,\phi_N(i)\,di\right] \\ &= \sum_{i=1}^{N} Var\left[\boldsymbol{v}(i)\,\phi_N(i)\right] di \\ &= \sum_{i=1}^{N} \Sigma_{vv}\phi_N(i)^2\,di \\ &= \zeta(N)\,\Sigma_{vv} \end{aligned}$$

where in moving to the second line we use the independence of each vector to ignore the covariance terms. The limiting variance as $N \to \infty$ is therefore $\zeta^*\Sigma_{vv}$ 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 (the Zipf distribution)

$$\phi_N(i) = c_N i^{-\gamma}; \text{ where } c_N = \left(\sum_{i=1}^{N} i^{-\gamma}\right)^{-1} \text{ and } \gamma > 1$$

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 1.2 plots the values of ζ^* for a range of values of γ for the Zeta distribution.[5]

This thesis explores two separate settings in which power law distributions are of economic importance. The first, studied in chapters 2 and 3 relates to *social*

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

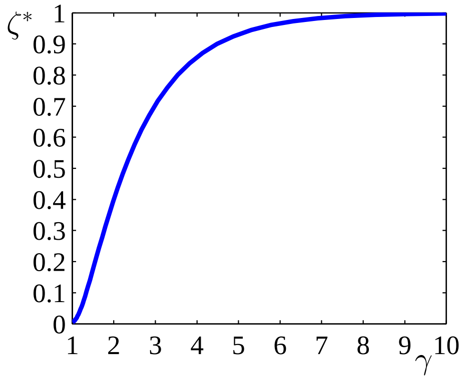


Figure 1.2: A plot of ζ^* for power law distributions with shape parameter γ

networks. A great many observed networks, from links between pages on Wikipedia to established relationships in social networks, have been shown to have degree distributions[6] well approximated by power law distributions (i.e. the networks are *scale free*). See, for example, the work of Albert and Barabási (2002), Jackson and Rogers (2007) or Clauset, Shalizi, and Newman (2009). The second setting, studied in chapter 4, relates to the distribution of *firm sizes*, which has also been shown to follow a power law. See Axtell (2001) or Gabaix (2011).

It is important to appreciate, though, that the models in this thesis do not generally assume any particular distribution, only that it remains non-uniform (in the sense of definition 4) as the support of that distribution grows arbitrarily large.

[6]In network theory, the *degree* of a node is the number of connections it has to other nodes, so the *degree distribution* is the distribution of these degrees over the entire network.

Chapter 2

Social learning over an opaque network

Abstract

I present a flexible and readily implemented linear model of rational (i.e. Bayesian) social learning over a network where agents do not know the full structure of the network, but do know the link distribution. I assume that there are several dynamic state variables to be estimated; agents act repeatedly and simultaneously; and agents' payoffs depend both on the accuracy of their beliefs regarding the state and the proximity of their actions to those of their competitors (i.e. there is strategic interaction). When the network is sufficiently irregular, transitory idiosyncratic shocks will not wash out in aggregation but will instead have persistent aggregate effects, and an AR(1) process for the underlying state will induce an ARMA(1,1) law of motion for the hierarchy of aggregate expectations.

2.1 Introduction

2.1.1 Context

An ideal model of network learning must contain a number of key features. Naturally, there must exist a hidden state of the world and agents be arranged in some sort of *observation network*, whereby they are informed of the actions, if not the actual beliefs, of a subset of their compatriots. Beyond this, there are arguably three desirable attributes of a "true" model of network learning:

1. agents act repeatedly;

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

The complexities involved in solving such a model, let alone simulating it or nesting it within a broader model of the economy, have typically been thought to be sufficiently great as to preclude comprehensive analysis in anything other than trivially small networks. As such, the literature to date has proceeded by avoiding one or more of the above assumptions.[1]

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, 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[2] 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,[3] they do not meet the

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

[2] 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 $[0, 1]$.

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

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 1.2 in the previous chapter 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[4] 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

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

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[5] and in which agents' payoffs depend only on the proximity of their expectation to the state (i.e. without 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.

2.1.2 This paper

In contrast to earlier work, the present paper is able to embrace all three of the assumptions listed above 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 (i.i.d.) distribution from which observation targets are drawn and do not learn about the structure of the network over time, agents' state vector of interest includes an infinite sequence of *weighted average* expectations instead of individual agents' expectations. Because of the recursive nature of agents' learning, this sequence will be of decreasing importance to the hierarchy of simple-average expectations, so an arbitrarily accurate approximation of the full solution may then be found by selecting a sufficiently high

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

cut-off for the number of weighted-average expectations 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, we are able to identify laws of motion for the agents' aggregate beliefs, even if we can never pin down the path of any individual's expectation.

The second requirement mentioned above – 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 and in such a scenario, agents are not *able* to learn about the network as it does not persist over time. The decisions followed by agents will then be fully rational in that they are entirely model-consistent, but the extent to which it might realistically be considered a "network" is called into question.[6] Second, 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.

With unobserved aggregate variables following an AR(1) process, we demonstrate that the full hierarchy of agents' aggregate expectations will follow an ARMA(1,1) process, with current and lagged weighted sums of agents' *idiosyncratic* shocks entering at an aggregate level. For sufficiently irregular networks – i.e. where some agents' actions are disproportionately observable – these weighted sums are shown to

[6] Such a setup, which is indeed deployed in this chapter, might arguably be better thought of as a search model.

not converge to zero, thereby adding aggregate volatility to the system. Despite idiosyncratic shocks being purely transitory, the aggregate volatility they induce through the network is also shown to exhibit (endogenous) persistence.

Because we examine a setting with a dynamic underlying state and demonstrate a specific law of motion for the hierarchy of average expectations, the researcher is able to simulate the aggregate effects of network learning[7] without a need to simulate individual agents' decisions. This makes the model particularly amenable to nesting within broad general equilibrium models of the economy.

Methodologically, this chapter expands on the work of Nimark (2008, 2011a,b), who in turn extended that of Woodford (2003). The latter of these reintroduced the ideas of Lucas (1972) and Phelps (1984) – that imperfect information could give rise to nominal shocks having real effects – and demonstrated that with incomplete information and strategic interaction, firms become interested in higher-order beliefs. In particular, with firms observing independent, unbiased signals of nominal GDP, Woodford demonstrated aggregate rigidity broadly equivalent to that produced by Calvo (1983) pricing. In contrast, Nimark (2008) supposed that firms' uncertainty surrounded the supply side of the economy (the average marginal cost) and, in addition to their private signals, granted firms visibility of the previous period's aggregate price level. Because aggregate variables are functions of the entire hierarchy of average expectations, this addition required the development of a new solution methodology that the present chapter adapts and extends to the idea of agents observing the previous-period actions of specific competitors.

Although the present chapter borrows from these papers and the next chapter applies the current model to a setting of firms' price-setting behaviour, the model developed here is context free and may be applied to any general setting with strategic competition and network learning. The conclusion considers a number of examples of such applications.

The remainder of this chapter is organised as follows. Section 2.2 presents the model, together with a characterisation of the solution and a methodology for finding

[7]The model is calibrated by two additional parameters: one specifying the number of other agents each player observes, and one describing the degree of irregularity in the network.

it. Section 2.3 provides an illustrative example of the model in action, applying it to a common scenario in the social learning literature. Section 2.4 considers some other applications of the model, including the need to consider dynamic actions (i.e. where agents' decision making includes consideration of competitors' future actions). Section 2.5 concludes.

2.2 The Model

We here develop a generalised model of Bayesian learning across opaque, stochastic networks with agents' optimal decision rule depending on both their expectation of the underlying state and their expectation of the agents' average action. The underlying state is assumed to be subject to persistent AR(1) shocks, while agents' observations include errors that are entirely transitory Gaussian white noise. It is demonstrated that for an irregular observation network, the hierarchy of aggregated expectations regarding the underlying state follows an ARMA(1,1) process, with weighted sums of agents' idiosyncratic shocks entering at the aggregate level (i.e. idiosyncratic shocks do not wash out).

A simple roadmap of how this section will proceed may be of some assistance. First, in subsection 2.2.1, we will describe the agents' problem, the information available to them and how they make their decisions. Subsection 2.2.3 will characterise agents' average action and briefly describe the informational assumptions used in previous research and how they differ to the current paper. Subsection 2.2.4 then explores the process of observing the actions of individual competitors before subsection 2.2.5 presents the main result of this paper.

2.2.1 The general setting

There is a countably infinite number of agents,[8] indexed in a continuum between zero and unity.[9] The *underlying state* follows a vector autoregressive process:

$$\boldsymbol{x}_t = A\boldsymbol{x}_{t-1} + P\boldsymbol{u}_t \tag{2.1}$$

where $\boldsymbol{u}_t$ is a vector of shocks with mean zero, while A and P are appropriately dimensioned matrices of fixed and publicly known parameters. Agents do not observe the value of $\boldsymbol{x}_t$ and must instead form beliefs about it. We define X_t to be the hierarchy of expectations regarding $\boldsymbol{x}_t$, in the sense of definition 3, and refer to it as

[8] 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.

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

the *state vector of interest.*

$$X_t \equiv \mathbb{E}_t^{(0:\infty)}\left[\boldsymbol{x}_t\right] \tag{2.2}$$

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

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

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

Agents' decision rule

Agents determine their individual actions simultaneously and according to a common linear decision rule:[10]

$$g_t\left(i\right) = \boldsymbol{\lambda}_1' E_t\left(i\right)\left[X_t\right] + \boldsymbol{\lambda}_2' \boldsymbol{x}_t + \boldsymbol{\lambda}_3' \boldsymbol{v}_t\left(i\right) \tag{2.3}$$

where $E_t\left(i\right)\left[\cdot\right] \equiv E\left[\cdot|\mathcal{I}_t\left(i\right)\right]$ is agent i's (first-order) expectation of the element within the square brackets conditional on all information available to her in period t (defined below); and $\boldsymbol{v}_t\left(i\right)$ 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 $\boldsymbol{v}_t\left(i\right)$ 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.[11]

[10]The derivation of the decision rule will invariably be context-specific. For example, in the context of price-setting to be explored in the next chapter, the underlying state will include aggregate shocks to marginal cost and demand, while agents' actions will be the price they choose and the signal they receive will include their private marginal cost and the previous-period price of a competitor.

[11]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.

Equation (2.3) is a reduced form expression for agents' actions that nests a wide array of commonly studied settings. An illustrative example with a univariate state is considered in depth in section 2.3 below, while other applications and their implied decision rules are discussed in section 2.4.

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:

$$\mathcal{I}_0(i) = \{\Omega, \Phi\} \qquad \mathcal{I}_t(i) = \{\mathcal{I}_{t-1}(i), \boldsymbol{s}_t(i)\} \tag{2.4}$$

where Ω is the set of all system coefficients, $\boldsymbol{s}_t(i)$ is the signal vector received each period and $\Phi : [0,1] \to [0,1]$ is the (cumulative) distribution from which agents' observation targets in the 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 – a *public/private signal* based on the underlying state and a *social signal* derived from observing competitors' actions with a one-period lag:

$$\boldsymbol{s}_t(i) = \begin{bmatrix} \boldsymbol{s}_t^p(i) \\ \boldsymbol{s}_t^s(i) \end{bmatrix} \tag{2.5}$$

$$\boldsymbol{s}_t^p(i) = D_1\boldsymbol{x}_t + D_2X_{t-1} + R_1\boldsymbol{v}_t(i) + R_2\boldsymbol{e}_t + R_3\boldsymbol{z}_{t-1}$$

$$\boldsymbol{s}_t^s(i) = \boldsymbol{g}_{t-1}(\boldsymbol{\delta}_{t-1}(i))$$

Public/Private signals may include both current and lagged information[12] 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.

[12] That is, they allow for some signals to only be observable with a one-period lag.

- $\boldsymbol{z}_t$ is a vector of *network shocks* (see equation 2.12 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}.

Note that 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 even formed them! However, we include terms in X_{t-1} and $\boldsymbol{z}_{t-1}$ (instead of just $\boldsymbol{x}_{t-1}$) to allow agents to observe, with a lag, aggregate variables and thus the past effect(s) 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:

$$\boldsymbol{\delta}_t : [0,1] \to [0,1]^q \tag{2.6}$$

where q is the number of agents observed. In other words, $\boldsymbol{\delta}_t\left(i\right)$ is the result of i's q separate draws from Φ for period t. For presentational simplicity, we will typically assume that $q = 1$ (i.e. that all agents observe a single other agent) and simply write $j = \delta_t\left(i\right)$ to mean that agent j's period-t action will be observed by agent i (in period $t+1$). To speak of the observee of an observee, we write $\delta_s\left(\delta_t\left(i\right)\right)$: the identity of the agent whose period-s action is observed by the agent whose period-t action is observed by agent i.

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

$$\begin{aligned} \boldsymbol{s}_t^s\left(i\right) &= g_{t-1}\left(\delta_{t-1}\left(i\right)\right) \\ &= \boldsymbol{\lambda}_1' E_{t-1}\left(\delta_{t-1}\left(i\right)\right)\left[X_{t-1}\right] + \boldsymbol{\lambda}_2' \boldsymbol{x}_{t-1} + \boldsymbol{\lambda}_3' \boldsymbol{v}_{t-1}\left(\delta_{t-1}\left(i\right)\right) \end{aligned} \tag{2.7}$$

2.2.2 The observation network

Because agent i's social signal is based on her observee's expectation, Bayesian updating then requires that i include $E_t\left(\delta_t\left(i\right)\right)\left[X_t\right]$ in her own state vector of interest. However, knowing that agent $\delta_t\left(i\right)$ is himself observing $\delta_t\left(\delta_t\left(i\right)\right)$ 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 1.2 of the previous chapter. In order to make the problem tractable, we 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, we 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 distribution is asymptotically non-uniform (see definition 4).*

As shown in the previous chapter and expanded on below, assumption 2 is sufficient to ensure that idiosyncratic shocks do not "wash out" in aggregation and will, in this context, enter into agents' aggregate beliefs. Social networks are widely regarded as having degree distributions well approximated by power law distributions,[13] which satisfy this assumption. It is important to appreciate, though, that the model here does not require any particular distribution of links in the agents' network, only that the network remain irregular as it grows arbitrarily large.

[13] See Albert and Barabási (2002), Jackson and Rogers (2007) or Clauset, Shalizi, and Newman (2009).

2.2.3 Agents' learning and imperfect common knowledge

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

$$X_t \equiv \mathbb{E}_t^{(0:\infty)}\left[\boldsymbol{x}_t\right] = FX_{t-1} + G_1\boldsymbol{u}_t + G_2\boldsymbol{z}_t + G_3\boldsymbol{e}_t + G_4\boldsymbol{z}_{t-1} \tag{2.8}$$

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 2.

In the macroeconomic literature, this environment – a state space system paired with strategic complementarity – is typically referred to as a setting of *incomplete common knowledge*, a phrase coined by Woodford (2003) in his demonstration of the potential importance of incomplete information in explaining inflation dynamics.

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 (2.5) and because of the moving average component of the law of motion (2.8). The usual response to these quirks would be 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 AR(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.

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

$$E_t\left(i\right)\left[\boldsymbol{z}_t\right] = \mathbf{0} \;\forall i, t \tag{2.9}$$

Proof. Since all shocks are distributed Normally, 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 $Cov\left(\boldsymbol{z}_t, \boldsymbol{s}_t\left(i\right)\right) = \mathbf{0}$. 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 $\boldsymbol{v}_t\left(i\right)$. However, we have that:

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

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

Since all 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[14] will be a Kalman filter:[15]

$$E_t\left(i\right)\left[X_t\right] = E_{t-1}\left(i\right)\left[X_t\right] + K\left\{\boldsymbol{s}_t\left(i\right) - E_{t-1}\left(i\right)\left[\boldsymbol{s}_t\left(i\right)\right]\right\} \tag{2.10}$$

where K is a time-invariant projection matrix (the Kalman gain). As in other models of imperfect common knowledge, since X_t includes $\overline{\boldsymbol{x}}_{t|t}^{(0:\infty)}$, we have 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.

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

[15] 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)).

Woodford (2003) supposed that agents (firms) each receive only a private signal regarding the underlying state (aggregate expenditure) and no social signal from other agents. In such a setting, where $X_t = \overline{\boldsymbol{x}}_{t|t}^{(0:\infty)}$, Woodford showed that F will be lower-triangular: each order of average expectations will be a linear combination of current period shocks and *lower* order expectations. Consequently, F may be constructed sequentially, first finding an expression for $E_t(i)[\boldsymbol{x}_t]$, then averaging it and repeating the process to find $E_t(i)\left[\overline{E}_t[\boldsymbol{x}_t]\right]$ and so forth.

By contrast, Nimark (2008) allowed agents to observe an aggregate signal (the average price) from the previous period in addition to their private signals.[16] This meant that each agent's signal vector includes a linear combination of the *entire hierarchy* of previous-period expectations (since individual actions are based on (expectations of) the entire hierarchy). 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 $X_t = \begin{bmatrix} \overline{\boldsymbol{x}}_{t|t}^{(0:\infty)\prime} & \overline{\boldsymbol{x}}_{t-1|t-1}^{(0:\infty)\prime} \end{bmatrix}'$.

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:

$$E_t(i)\left[\overline{\boldsymbol{x}}_{t|t}^{(0:\infty)}\right] = K\boldsymbol{s}_t(i) + (F - K(D_1F + D_2))\,E_{t-1}(i)\left[\overline{\boldsymbol{x}}_{t-1|t-1}^{(0:\infty)}\right]$$

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

It is perhaps worth emphasising that the signal structures assumed by Woodford (2003) and Nimark (2008) both result in agents only being concerned with the *simple average* expectation of their peers (or higher-order versions of the same). In the language of chapter 1, they have chosen signal structures that explicitly set $p = 1$, thereby having the infinite dimensionality of the state vector arising only from the presence of higher-order expectations.

[16] The hidden aggregate state in Woodford (2003) was on the demand side of the economy, while that in Nimark (2008) was on the supply side. Woodford also limited attention to static pricing by firms, while Nimark made use of Calvo-style dynamic pricing.

2.2.4 Observing individual competitors' actions

In implementing the Kalman filter (2.10), agent i needs to create a prior expectation of the signal they will receive in the next period. From equation (2.7), we see that it is therefore necessary for agent i to construct $E_t(i)[g_t(\delta_t(i))]$ as part of her prior for period $t+1$, which includes $E_t(i)[E_t(\delta_t(i))[X_t]]$:

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

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 $\delta_{t-1}(\delta_t(i))$.

Proposition 1. *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.*

Proof. The proof may be found in appendix 2.A. □

We see from equation (2.3) that the weighted-average action, $\widetilde{g}_t$, is given by:

$$\widetilde{g}_t = \boldsymbol{\lambda}_1'\widetilde{E}_t[X_t] + \boldsymbol{\lambda}_2'\boldsymbol{x}_t + \boldsymbol{\lambda}_3'\widetilde{\boldsymbol{v}}_t \tag{2.11}$$

where $\widetilde{E}_t[\cdot] \equiv \int_0^1 E_t(j)[\cdot]\,\phi(j)\,dj$ is the (first-order) weighted-average expectation, which we will more fully denote ${}^{\{1\}}\widetilde{E}_t[\cdot]$ (see proposition 2 below).

We cannot, in general, make use of some law of large numbers to disregard the effect of idiosyncratic shocks in the weighted-average action – that is, we 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 $\phi(i) = 0\ \forall i \neq 1$), we would then have that $\widetilde{\boldsymbol{v}}_t = \boldsymbol{v}_t(1)$ which in any given period will be non-zero, almost surely.

In this regard, assumptions 1 and 2 are sufficient to allow us to assert the following proposition regarding the limiting properties of aggregate (random) variables derived from agents' idiosyncratic shocks:

Proposition 2. *Suppose that* $\boldsymbol{v}_t(i) \sim i.i.d.\ N(\mathbf{0}, \Sigma_{vv})\ \forall i, t$. *For a finite number of agents* (N), *define the aggregate statistics*

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

Given assumptions 1 and 2, we have the following results in the limit (as $N \to \infty$*):*

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

Proof. The proof may be found in appendix 2.B. □

Given proposition 2, we refer to ${}^{\{q\}}\widetilde{\boldsymbol{v}}_t$ as the ***q*-th weighted average** of agents' idiosyncratic shocks and define the vector of **network shocks**, $\boldsymbol{z}_t$, as that containing the full sequence of these weighted sums:

$$
\boldsymbol{z}_t \equiv \begin{bmatrix} {}^{\{1\}}\widetilde{\boldsymbol{v}}_t \\ {}^{\{2\}}\widetilde{\boldsymbol{v}}_t \\ {}^{\{3\}}\widetilde{\boldsymbol{v}}_t \\ {}^{\{4\}}\widetilde{\boldsymbol{v}}_t \\ \vdots \end{bmatrix} \sim N(\mathbf{0}, \Sigma_{zz}) \qquad \Sigma_{zz} = \begin{bmatrix} \Sigma^{\{1\}}_{\widetilde{v}\widetilde{v}} & \Sigma^{\{1\}}_{\widetilde{v}\widetilde{v}} & \Sigma^{\{1\}}_{\widetilde{v}\widetilde{v}} & \Sigma^{\{1\}}_{\widetilde{v}\widetilde{v}} & \cdots \\ \Sigma^{\{1\}}_{\widetilde{v}\widetilde{v}} & \Sigma^{\{2\}}_{\widetilde{v}\widetilde{v}} & \Sigma^{\{2\}}_{\widetilde{v}\widetilde{v}} & \Sigma^{\{2\}}_{\widetilde{v}\widetilde{v}} & \cdots \\ \Sigma^{\{1\}}_{\widetilde{v}\widetilde{v}} & \Sigma^{\{2\}}_{\widetilde{v}\widetilde{v}} & \Sigma^{\{3\}}_{\widetilde{v}\widetilde{v}} & \Sigma^{\{3\}}_{\widetilde{v}\widetilde{v}} & \cdots \\ \Sigma^{\{1\}}_{\widetilde{v}\widetilde{v}} & \Sigma^{\{2\}}_{\widetilde{v}\widetilde{v}} & \Sigma^{\{3\}}_{\widetilde{v}\widetilde{v}} & \Sigma^{\{4\}}_{\widetilde{v}\widetilde{v}} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{2.12}
$$

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 the aggregate expectation ${}^{\{1\}}\widetilde{E}_t[X_t]$ will be a function of ${}^{\{1\}}\widetilde{\boldsymbol{v}}_t$ and ${}^{\{2\}}\widetilde{E}_{t-1}[X_{t-1}]$, the latter of which will be a function of ${}^{\{2\}}\widetilde{\boldsymbol{v}}_{t-1}$ and ${}^{\{3\}}\widetilde{E}_{t-2}[X_{t-2}]$, etc.

Note that the following two corollaries immediately follow from proposition 2:

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

Proof. Trivial, since $\zeta^* \in (0, 1)$. □

Corollary 2. $E\left[{}^{\{q\}}\widetilde{v}_t \,|\, {}^{\{1\}}\widetilde{v}_t = \boldsymbol{a}\right] = \boldsymbol{a}\ \forall q \geq 2$

Proof. Follows immediately from item 3 in the proposition. □

The first of these is a necessary component of approximating the full solution with a finite state vector (see section 2.2.6 below), while the latter is used when simulating the aggregate effects of network learning.

2.2.5 Social learning over an opaque, irregular network

We are now in a position to present the main result of this chapter.

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

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

where

$$\begin{aligned} \overline{E}_t\left[\cdot\right] &= \int_0^1 E_t\left(i\right)\left[\cdot\right] di \\ {}^{\{1\}}\widetilde{E}_t\left[\cdot\right] &= \int_0^1 E_t\left(\delta_t\left(i\right)\right)\left[\cdot\right] di \\ {}^{\{2\}}\widetilde{E}_t\left[\cdot\right] &= \int_0^1 E_t\left(\delta_t\left(\delta_t\left(i\right)\right)\right)\left[\cdot\right] di \\ &\vdots \end{aligned}$$

Proof. The proof may be found in appendix 2.C □

Although the complete derivation is provided in the appendix, an outline of the agents' learning process may be of interest. To begin, we define the matrices S_x, T_s and T_{w_q} as the matrices that select $\boldsymbol{x}_t$, $\overline{E}_t[X_t]$ and ${}^{\{q\}}\widetilde{E}_t[X_t]$ respectively from X_t (e.g., $T_{w_2}X_t = {}^{\{2\}}\widetilde{E}_t[X_t]$). We also define the general notation that $\theta_{t|q}(i)$ represents the *error* in agent i's period-q expectation regarding θ_t. In particular, we will use the following:

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

The filter

As with a standard Kalman filter, the Kalman gain (equation 2.30) is calculated as:

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

where $\boldsymbol{s}_{t|t-1}(i)$ is the agent's signal innovation (the portion of their signal that was not forecastable). With agents observing the previous-period actions of specific competitors, the signal innovation is then able to be expressed (equation 2.36) as:

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

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 error. As such, both the covariance and variance terms in the Kalman gain will therefore include terms in both the variance of i's expectation error, $V_{t-1|t-1} \equiv E\left[X_{t-1|t-1}(i)\, X_{t-1|t-1}(i)'\right]$, and the *covariance* between any two agents' errors, $W_{t-1|t-1} \equiv E\left[X_{t-1|t-1}(i)\, X_{t-1|t-1}(j)'\right]$.

The variance in agents' expectation errors then updates in the usual way via an interim prior variance, but a corresponding expression must also be found for updating the covariance between agents' errors.

Defining $M \equiv \begin{bmatrix} M_1 & M_2 & M_3 \end{bmatrix}$, the full set of equations for updating the filter for one period is given by:

$$\begin{aligned} E\left[\boldsymbol{s}_{t|t-1}(i)\,\boldsymbol{s}_{t|t-1}(i)'\right] = M &\begin{bmatrix} V_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \\ W_{t-1|t-1} & V_{t-1|t-1} & V_{t-1|t-1} \\ V_{t-1|t-1} & V_{t-1|t-1} & U_{t-1} \end{bmatrix} M' \\ &+ (M_1 + M_2 + M_3)\, G_2 \Sigma_{zz} N_5' \\ &+ N_5 \Sigma_{zz} G_2' (M_1 + M_2 + M_3)' \\ &- M_2 K_{t-1} N_2 \Sigma_{vv} N_4' \\ &- N_4 \Sigma_{vv} N_2' K_{t-1}' M_2' \\ &+ N_1 \Sigma_{uu} N_1' + N_2 \Sigma_{vv} N_2' + N_4 \Sigma_{vv} N_4' \end{aligned} \tag{2.13a}$$

$$\begin{aligned} E\left[\boldsymbol{s}_{t|t-1}(i)\,\boldsymbol{s}_{t|t-1}(j)'\right] = M &\begin{bmatrix} W_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \\ W_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \\ V_{t-1|t-1} & V_{t-1|t-1} & U_{t-1} \end{bmatrix} M' \\ &+ (M_1 + M_2 + M_3)\, G_2 \Sigma_{zz} N_5' \\ &+ N_5 \Sigma_{zz} G_2' (M_1 + M_2 + M_3)' \\ &+ N_1 \Sigma_{uu} N_1' \end{aligned} \tag{2.13b}$$

$$\begin{aligned} E\left[X_t \boldsymbol{s}_{t|t-1}(i)'\right] = F &\begin{bmatrix} V_{t-1|t-1} & V_{t-1|t-1} & U_{t-1} \end{bmatrix} M' \\ &+ G_1 \Sigma_{uu} N_1' \\ &+ F G_2 \Sigma_{zz} N_5' \\ &+ G_4 \Sigma_{zz} G_2' (M_1 + M_2 + M_3)' \\ &+ G_4 \Sigma_{zz} N_5' \end{aligned} \tag{2.13c}$$

$$\begin{aligned} E\left[X_{t|t-1}(i)\,\boldsymbol{s}_{t|t-1}(j)'\right] = F &\begin{bmatrix} V_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \end{bmatrix} M' \\ &+ G_1 \Sigma_{uu} N_1' \\ &+ F G_2 \Sigma_{zz} N_5' \\ &+ G_4 \Sigma_{zz} G_2' (M_1 + M_2 + M_3)' \\ &+ G_4 \Sigma_{zz} N_5' \end{aligned} \tag{2.13d}$$

$$K_t = E\left[X_t \boldsymbol{s}_{t|t-1}(i)'\right] \left(E\left[\boldsymbol{s}_{t|t-1}(i)\, \boldsymbol{s}_{t|t-1}(i)'\right]\right)^{-1} \tag{2.13e}$$

$$\begin{aligned} U_t &= F U_{t-1} F' \\ &\quad + G_1 \Sigma_{uu} G_1' + G_2 \Sigma_{zz} G_2' + G_4 \Sigma_{zz} G_4' + F G_2 \Sigma_{zz} G_4' + G_4 \Sigma_{zz} G_2' F' \end{aligned} \tag{2.13f}$$

$$\begin{aligned} V_{t|t-1} &= F V_{t-1|t-1} F' \\ &\quad + G_1 \Sigma_{uu} G_1' + G_2 \Sigma_{zz} G_2' + G_4 \Sigma_{zz} G_4' + F G_2 \Sigma_{zz} G_4' + G_4 \Sigma_{zz} G_2' F' \end{aligned} \tag{2.13g}$$

$$\begin{aligned} W_{t|t-1} &= F W_{t-1|t-1} F' \\ &\quad + G_1 \Sigma_{uu} G_1' + G_2 \Sigma_{zz} G_2' + G_4 \Sigma_{zz} G_4' + F G_2 \Sigma_{zz} G_4' + G_4 \Sigma_{zz} G_2' F' \end{aligned} \tag{2.13h}$$

$$V_{t|t} = V_{t|t-1} - K_t\, E\left[\boldsymbol{s}_{t|t-1}(i)\, \boldsymbol{s}_{t|t-1}(i)'\right] K_t' \tag{2.13i}$$

$$\begin{aligned} W_{t|t} = W_{t|t-1} &+ K_t E\left[\boldsymbol{s}_{t|t-1}(i)\, \boldsymbol{s}_{t|t-1}(j)'\right] K_t' \\ &- E\left[X_{t|t-1}(i)\, \boldsymbol{s}_{t|t-1}(j)'\right] K_t' \\ &- K_t E\left[\boldsymbol{s}_{t|t-1}(i)\, X_{t|t-1}(j)'\right] \end{aligned} \tag{2.13j}$$

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

The law of motion

Starting from the basic form of the Kalman filter:

$$E_t(i)[X_t] = F E_{t-1}(i)[X_{t-1}] + K_t \boldsymbol{s}_{t|t-1}(i)$$

we substitute in the above expression for $\boldsymbol{s}_{t|t-1}(i)$ and take a simple average 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 ${}^{\{1\}}\widetilde{E}_{t-1}[X_{t-1}]$, thereby introducing the need to also determine the (first) weighted-average expectation.

Taking the weighted average of the filter to obtain ${}^{\{1\}}\widetilde{E}_t\left[X_t\right]$ then produces a term in ${}^{\{2\}}\widetilde{E}_{t-1}\left[X_{t-1}\right]$, thus requiring that we include the second weighted average expectation. The second weighted average expectation subsequently produces a term in the third weighted average expectation, and so forth.

The coefficients for the full-state law of motion are given by:

$$F = \begin{bmatrix} \begin{bmatrix} A & \mathbf{0}_{m\times\infty} \end{bmatrix} \\ K\left(M_1 + M_2 + M_3\right) + \left(F - KM_1\right)T_s - KM_2T_{w_1} \\ K\left(M_1 + M_2 + M_3\right) + \left(F - KM_1\right)T_{w_1} - KM_2T_{w_2} \\ K\left(M_1 + M_2 + M_3\right) + \left(F - KM_1\right)T_{w_2} - KM_2T_{w_3} \\ \vdots \end{bmatrix} \tag{2.14a}$$

$$G_1 = \begin{bmatrix} P \\ KN_1 \\ KN_1 \\ KN_1 \\ \vdots \end{bmatrix} \quad G_2 = \begin{bmatrix} \mathbf{0}_{m\times\infty} \\ \mathbf{0}_{\infty\times\infty} \\ K\begin{bmatrix} N_2 & \mathbf{0}_{1\times r} & \mathbf{0}_{1\times r} & \mathbf{0}_{1\times\infty} \end{bmatrix} \\ K\begin{bmatrix} \mathbf{0}_{1\times r} & N_2 & \mathbf{0}_{1\times r} & \mathbf{0}_{1\times\infty} \end{bmatrix} \\ \vdots \end{bmatrix} \tag{2.14b}$$

$$G_3 = \begin{bmatrix} \mathbf{0}_{m\times n} \\ KN_3 \\ KN_3 \\ KN_3 \\ \vdots \end{bmatrix} \quad G_4 = \begin{bmatrix} \mathbf{0}_{m\times\infty} \\ K\left(\begin{bmatrix} N_4 & \mathbf{0}_{1\times r} & \mathbf{0}_{1\times r} & \mathbf{0}_{1\times\infty} \end{bmatrix} + N_5\right) \\ K\left(\begin{bmatrix} \mathbf{0}_{1\times r} & N_4 & \mathbf{0}_{1\times r} & \mathbf{0}_{1\times\infty} \end{bmatrix} + N_5\right) \\ K\left(\begin{bmatrix} \mathbf{0}_{1\times r} & \mathbf{0}_{1\times r} & N_4 & \mathbf{0}_{1\times\infty} \end{bmatrix} + N_5\right) \\ \vdots \end{bmatrix} \tag{2.14c}$$

where m is the number of elements in the underlying state ($\boldsymbol{x}_t$); n is the number of elements in the vector of public signal noise ($\boldsymbol{e}_t$); and r is the number of elements in each agents' vector of idiosyncratic shocks ($\boldsymbol{v}_t\left(i\right)$).

Since these matrices are defined recursively, finding the solution involves finding the fixed point of the system for a given Kalman gain (K).

2.2.6 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, so the full solution cannot be found in practice.

For standard problems with imperfect common knowledge, where only the hierarchy of simple-average expectations is needed,[17] 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 by agents to higher-order average expectations is decreasing in the order of expectation; 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 $\boldsymbol{\lambda}_1$.[18] 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 we can write $X_t = E_t(j)[X_t] + X_{t|t}(j)$ and the variance of the two sides must be equal, we have

$$Var(X_t) = Var(E_t(j)[X_t]) + Var\left(X_{t|t}(j)\right)$$

where we can ignore the covariance term on the right hand side 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

[17] That is, where there is only one compound expectation of interest ($p = 1$).
[18] See section 2.3 for a typical example.

the weighted-average of those shocks will be less the variance of an individual shock as shown above in corollary 1 to proposition 2. We therefore have that

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

The recursive structure of X_t then establishes the result.

With network learning over an opaque network, however, it is *also* necessary 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 by agents to higher-weighted average 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 2.

2.2.7 Finding the solution

The solution is defined implicitly in two respects (the filter and the law of motion), both of which require iterating through a series of update rules while taking the other as given.

Note that the size of the state vector can still be very large even when operating with few state variables (m) and quite low choices of k^* and p^*. Table 2.1 lists the sizes that emerge for a variety of parameters.

Given the size of the matrices involved, problems of *numerical instability* must be considered. Numerical instability arises as a consequence of the round-off errors that necessarily occur with floating-point operations on computers. When iterating a large

m	k^*	No network (standard ICK)	With network learning ($p^* = 3$)
1	4	5×5 : 200 Bytes	121×121 : 114.4 KB
1	6	7×7 : 392 Bytes	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 2.1: Size (each) of F, U, V and W, assuming use of double-precision.

system over many steps, these errors can accumulate and magnify to the extent that the system does not converge.[19] 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.

A variety of approaches can be undertaken to combat numerical instability. We list the major ones here, together with a description of how (if possible) each has been incorporated into the attached Matlab code.

Minimise the size of the state vector

The primary approach to avoiding numerical stability issues is, where possible, to reduce the size of the system being estimated. It is for this reason that the solution developed here makes use of Nimark (2011b) in avoiding the need to double the state vector (which would multiply the number of operations in a matrix multiplication by eight) when including lagged signals.

Factor the variance and covariance matrices

Arguably the most robust (to roundoff error) implementations of Kalman filters are those that factor the relevant variance-covariance matrices. In particular, since the variance matrices are by definition symmetric and positive (semi) definite, a Cholesky decomposition of them (i.e. the decomposition of V into LL' with L lower triangular) can be deployed. By operating on the L matrices directly, the implied variance matrices remain well defined.

[19]The number of arithmetic operations involved in matrix multiplication or inversion typically increases with the cube of the matrix's dimension, with roundoff errors able to enter in every operation

In practice, a modified Cholesky decomposition, sometimes referred to as a "UD decomposition", that breaks V into UDU' with U unit upper triangular (i.e. with ones on the leading diagonal) and D diagonal is typically used,[20] as this avoids the need to find any square roots. 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 ($\boldsymbol{s}_{t|t-1}(i)$) can no longer be expressed in the form

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

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 which accounted for the more complex structure of the Kalman gain found here. Such an investigation is left for future research.

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 (e.g., we might nest the finding of a time-invariant filter within each iteration of updating the law of motion), but such an algorithm is needlessly complex and in practice is more likely to suffer from numerical stability issues.

Instead, for a given set of signals, we find the fixed point by updating the filter and the law of motion incrementally within the same loop:

[20] Of course, an equivalent expression of $L^*DL^{*\prime}$ may be found, with L^* unit lower triangular.

repeat
  Update the filter by one step using equation (2.13)
  Update the law of motion by one step using equation (2.14)
until both the filter and the law of motion converge

Avoid temporary creation of unnecessarily large matrices

The solution as presented in the text (see equations 2.13a and 2.13b) 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 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.

Optimise the selection of initial conditions

Choosing initial values for F, U, V and W that are in some sense close to their final values has three benefits:

1. By lowering the number of iterations required, it reduces the time taken to converge to the final result;
2. Lowering the number of iterations (quite dramatically) lowers the number of calculations required, thereby reducing the opportunity for roundoff errors to affect the result; and

3. Starting closer to the final result reduces the chance of the agents' error variance needing to "pass through infinity" *en route* to the solution.

The latter point derives from the fact that sufficiently low initial variance will cause the system to diverge, while arbitrarily large initial variance will, in principle, converge asymptotically to the final result (see section 4.8.4.4 of Grewel and Andrews, 2008). Given this, a common practice is to impose exceptionally high initial variances in the hope of avoiding divergence. However, such a practice is also fraught in that too large an initial variance can be a particular source of numerical roundoff if it is sufficiently large relative to $Cov\left(X_t, \boldsymbol{s}_{t|t-1}\left(i\right)\right)$.

Instead, we make use of our assumption that agents' problems place decreasing weight against higher-order expectations to note that the natural set of initial conditions to choose when considering a problem with k^* orders of expectation is the solution to the problem with $(k^* - 1)$ orders.

We therefore use the following process to find the solution:

1. Starting at time $t = 0$ with $k^* = 2$,[21] suppose that agents observe no signals at all. This implies that the variance $(V_{t|t})$ and covariance $(W_{t|t})$ of agents' expectation errors will be equal to the unconditional variance of the full state (U_t).
2. Starting at the convergent result from step one, suppose that agents now observe their public and private signals. The no-network solution (for $k^* = 2$) is found via the algorithm listed above by imposing that $q = 0$.
3. Starting at the convergent result from step two, suppose that agents now observe all of their signals. The network learning solution for $k^* = 2$ is found via the algorithm listed above.
4. Starting at the convergent result from the previous step, increase k^* and solve the problem with agents observing both public/private and social signals.

[21] Note that solving the network learning model requires that $k^* \geq 2$, as the law of motion gives us that each of the compound expectations of order k relies on expectations of order $k - 2$.

The final step is repeated for higher and higher values of k^* until the researcher is satisfied that the results are a sufficiently accurate approximation of the $k^* \to \infty$ solution.

2.2.8 A special case

Recall that agents' decision rule is assumed to be given by

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

and their signal vector by

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

A special case emerges when agents' actions only depend on their beliefs regarding the hierarchy of simple-average expectations (so coefficients in $\boldsymbol{\lambda}_1$ against other compound expectations are all zero) and agents' signals from the previous period are based only on the hierarchy of simple-average expectations (so coefficients in D_2 against other compound expectations are all zero). In this case, we posit the following conjecture:

Conjecture 1. *When agents observe each others' individual actions in an opaque network and therefore attempt to estimate higher weighted average expectations, but this is done solely as part of estimating $\overline{\boldsymbol{x}}_{t|t}^{(0:\infty)}$ (i.e. individual agents' estimates of ${}^{\{q\}}\widetilde{E}_t[X_t]$ are of no direct economic significance to them) and agents observe lagged signals that are based only on $\overline{\boldsymbol{x}}_{t-1|t-1}^{(0:\infty)}$, then the impulse responses of $\overline{\boldsymbol{x}}_{t|t}^{(0:\infty)}$ following any aggregate or network shock are the same for any $p^* \geq 2$.*

In other words, for the purposes of simulating the effects of network learning, it is only necessary to include the primary compound expectation (here the simple average) and the first weighted average expectation from the network. We have not yet been able to prove this conjecture, but have confirmed that it holds for $p^* \in \{2, 3, 4\}$ with $k^* = 6$ and $m = 1$. Its proof clearly calls for future work.

2.3 An illustrative example

We here present a simplified example to illustrate some of the results that emerge from adding network learning to a setting of strategic complementarity. A more extensive model in the context of firms' price-setting decisions is presented in the next chapter.

2.3.1 The simplified model

There exists only a single hidden state that follows an AR(1) process

$$x_t = \rho x_{t-1} + u_t \qquad u_t \sim N\left(0, \sigma_u^2\right) \tag{2.15}$$

Agents each receive a single, unbiased private signal about the state

$$s_t^p\left(i\right) = x_t + v_t\left(i\right) \qquad v_t\left(i\right) \sim N\left(0, \sigma_v^2\right) \tag{2.16}$$

with u_t and $v_t\left(i\right)$ 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:[22]

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

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

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

Taking the average of this expression and repeated substituting it back in then eventually yields

$$g_t\left(i\right) = \left(1-\beta\right)\begin{bmatrix}1 & \beta & \beta^2 & \cdots\end{bmatrix} E_t\left(i\right)\left[\overline{x}_t^{(0:\infty)}\right] \tag{2.17}$$

[22]This utility function is quite common in the network literature. See, for example, 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).

In the framework presented above, this would be captured by

$$\boldsymbol{\lambda}_0 = \boldsymbol{\lambda}_2 = \boldsymbol{\lambda}_3 = \mathbf{0}$$
$$\boldsymbol{\lambda}_1 = (1-\beta)\begin{bmatrix}1 & \beta & \beta^2 & \cdots\end{bmatrix}(S_x + T_s)$$

Note that equations (2.16) and (2.17) satisfy the conditions for the special case laid out above in section 2.2.8. As a baseline, we suppose the following parameters:

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 2.2: Baseline parameterisation

2.3.2 Aggregate beliefs following a shock to the underlying state

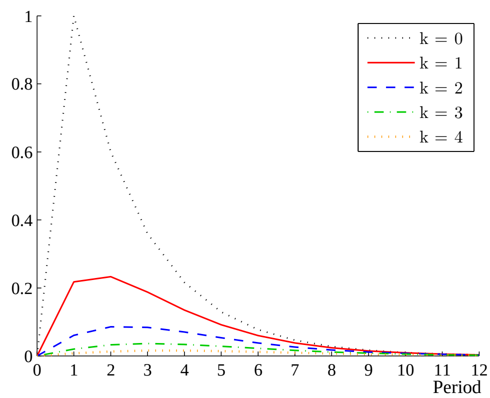


Figure 2.1: The hierarchy of simple-average expectations $(\overline{x}_{t|t}^{(0:k^*)})$ following a one standard deviation shock to the underlying state with no network $(q = 0)$

Figure 2.1 plots a standard scenario in the incomplete common knowledge literature, showing impulse responses for the resultant hierarchy of simple-average expectations[23] following a one standard deviation shock to the hidden state when there

[23]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.

is no network learning, so agents only have 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. Note that all orders of expectation remain below the the underlying state, so the average expectation error $(x_t - \overline{E}_t[x_t])$ remains strictly positive. The hierarchy of beliefs subsequently decays back to zero with the underlying shock.

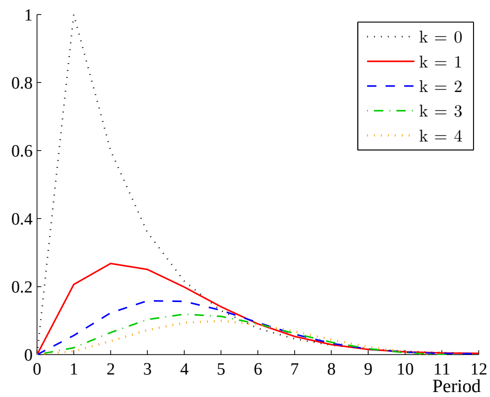
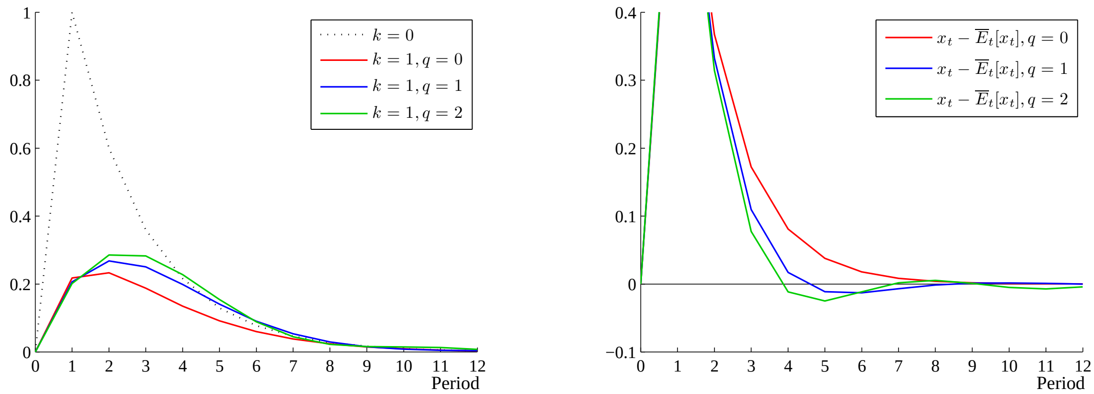


Figure 2.2: The hierarchy of simple-average expectations $(\overline{x}_{t|t}^{(0:k^*)})$ following a one standard deviation shock to the underlying state with agents each observing one competitor $(q = 1)$.

Figure 2.2 next plots the same hierarchy when, in addition to observing their private signals, each agent observes the previous-period action of one competitor. 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, the presence of network learning introduces a degree of persistence in agents' aggregate beliefs beyond that embodied in the underlying state so that agents' average expectations are *above* the truth (the average expectation error $(x_t - \overline{E}_t[x_t])$ becomes and remains strictly 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.[24] 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.

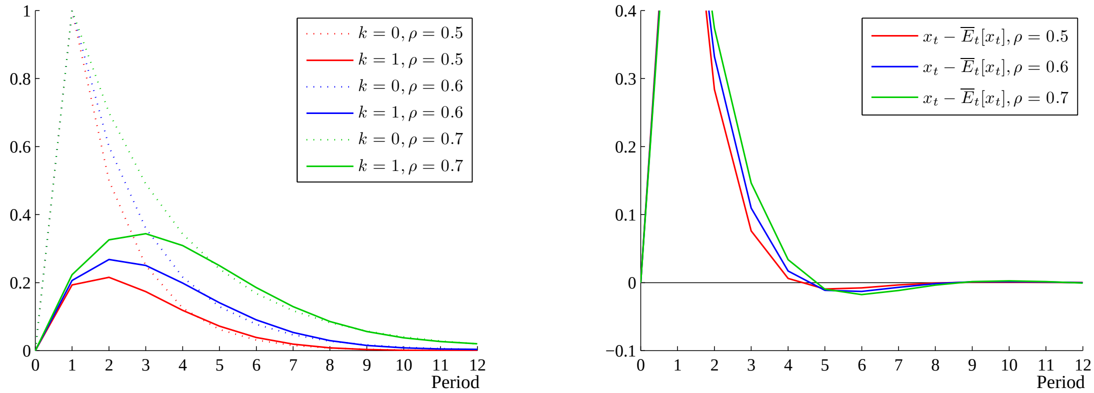


(a) Simple-average expectations (b) Average expectation errors

Figure 2.3: Varying the number of other agents observed (q)

Figure 2.3 illustrates the responses for different numbers of other agents observed. Although the addition of social signals lessens the average expectation error in the near term, in the longer term expectations overshoot so that, on average, errors become negative. The absolute value of the long-term average error is increasing in the number of competitors observed.

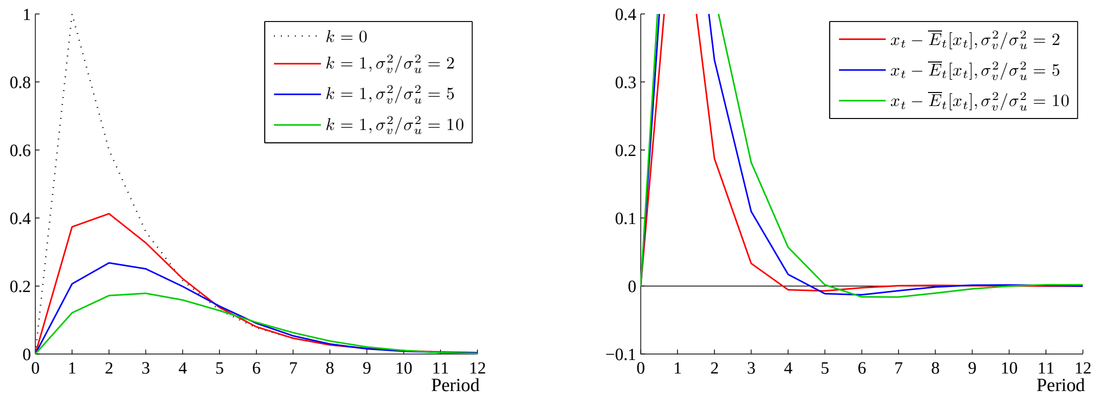
[24]Strictly speaking, agents do not assume that their competitors observe a public signal. Rather, their Bayes-rational signal extraction problem is mathematically equivalent to making the assumption.



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

Figure 2.4: Varying underlying persistence (ρ)

Figure 2.4 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 2.5: Varying the relative innovation variance (σ_v^2/σ_u^2)

Figure 2.5 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,[25] worsens the value of agents' private signals, causing them to rely more heavily on the social signals and so causing marginally worse performance in the longer term.

2.3.3 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 suite of idiosyncratic shocks in a period for which more prominent agents happen to draw innovations in one direction (say, positive) while more obscure agents draw innovations in the opposite direction. Overall, with a continuum of agents, the law of large numbers ensures that the simple average innovation is zero, but an average weighted by the agents' probability of being observed will be non-zero.

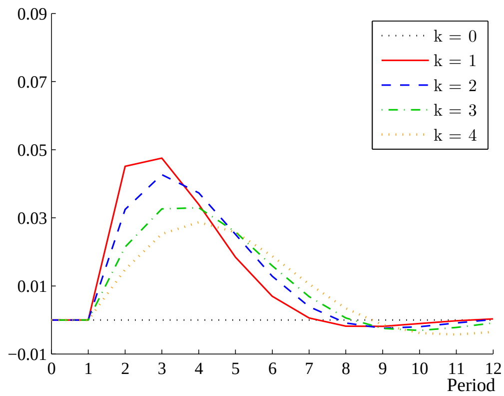


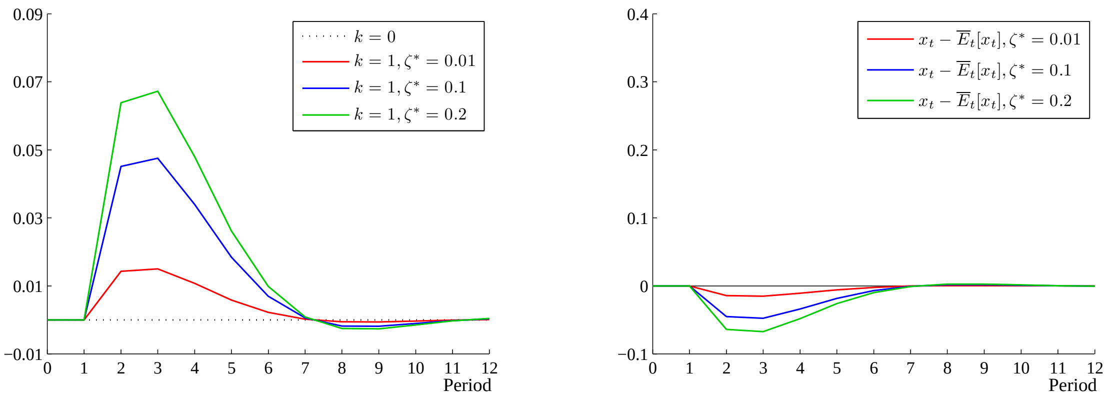
Figure 2.6: The hierarchy of simple-average expectations ($\overline{x}_{t|t}^{(0:k^*)}$) following a one standard deviation network shock (a one standard deviation shock to $\widetilde{v}_t$ and the corresponding conditional expected value for higher-weighted averages) with agents each observing one competitor ($q = 1$).

Figure 2.6 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 $\widetilde{v}_t$ plus the corresponding (conditionally) expected value for higher weighted averages – when agents each observe one competitor ($q = 1$). Note that the

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

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 for a univariate private signal, $Var\left(\widetilde{v}_t\right) = \zeta^* \sigma_v^2$, where $\zeta^* \in (0, 1)$ indicates the degree of irregularity in the network. Increasing the irregularity of the network (i.e. making the distribution of inbound observation links less uniform) therefore increases the scale of typical network shocks.



(a) Simple-average expectations

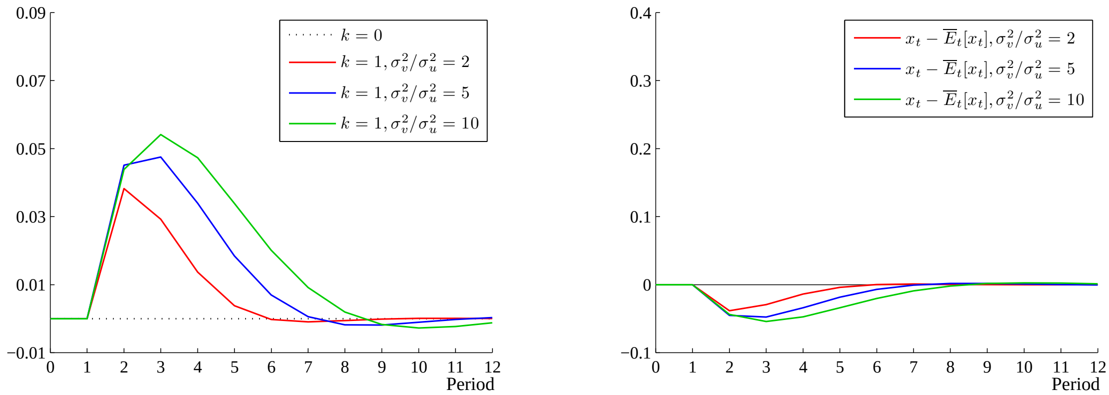
(b) Average expectation errors

Figure 2.7: Varying the degree of network irregularity (ζ^*)

Figure 2.7 illustrates this, plotting the first-order simple-average expectations and average expectation errors for a variety of values for ζ^* following a one standard deviation network shock when agents each observe a single competitor.

At one extreme, the network is regular ($\zeta^* \to 0$), so the distribution of links is uniform and the law of large numbers therefore 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. For the baseline scenario listed above, a one standard deviation network shock when $\zeta^* = 1$ produces a peak average expectation of 0.13, compared to the 0.31 obtained from a one standard deviation shock to the underlying state.

Note, too, that although varying ζ^* changes the magnitude of the movement in agents' average expectations, the persistence of that movement is unchanged across different values of ζ^*.



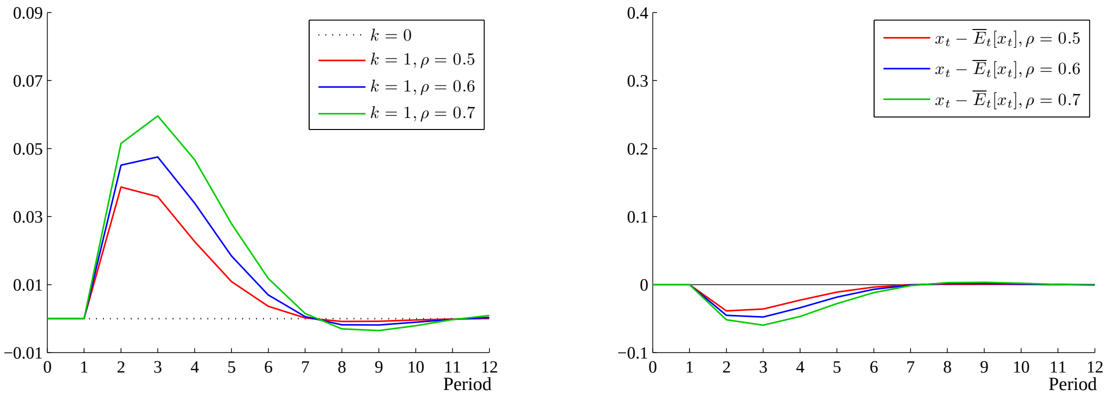
(a) Simple-average expectations

(b) Average expectation errors

Figure 2.8: Varying the relative innovation variance (σ_v^2/σ_u^2)

Figure 2.8 next shows the effect on network shocks from 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 shock's effects.

Finally, figure 2.9 shows that both the magnitude and the persistence of deviations in average expectations following a network shock increase with the persistence of



(a) Simple-average expectations

(b) Average expectation errors

Figure 2.9: Varying underlying persistence (ρ)

underlying system.

2.4 Other examples

The model developed here is arguably applicable to a number of areas of ongoing macroeconomic research. Possible applications may include

1. In a setting of monopolistic competition with firms' facing demand curves that are functions of their *relative* prices, price-setting firms may inform their decisions by observing the prices of individual competitors. Coordination concerns and network learning will then induce notably different dynamics for aggregate inflation. This setting is explored in detail in chapter 3 of this thesis.
2. When posting vacancies in a labour search model in the style of Mortensen and Pissarides (1994), firms' probability of finding a successful match is dependent on the number of vacancies that other firms post. When firms' productivity includes both aggregate and idiosyncratic components, observing the number of vacancies posted by their competitors allows firms to predict both the component of their productivity that is common to all and their expected gain from posting an additional vacancy themselves.
3. In the asset pricing model of Singleton (1987), traders' individual demand for a risky asset is dependent on their expectation of the next-period price, itself a function of all traders' actions and (unobserved in advance) shocks to the supply of the asset. Observing the actions of (some of) their competitors would allow traders to learn about the (higher-order) expectations of other traders and adjust their responses accordingly.

Three features arguably common to all of these possible applications are that (a) agents' actions are affected directly by their private signals *in addition* to indirectly through their expectations; (b) agents may need to consider the actions of other agents not just in the current period but also into the future; and (c) agents may receive public signals about aggregate statistics.

The broad model of this chapter readily nests all of these features. For 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:

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

We show in appendix 2.D 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 (2.3).

This is by no means the only dynamic setting that may be modelled here. The dynamic price-setting model explored in chapter 3, for example, considers an environment with an infinite sum of forward-looking variables in the individual firm's decision rule and the addition of a lagged public signal (the previous period's aggregate price).

2.5 Conclusion

This chapter has introduced and solved a model of social learning with a continuum of agents that satisfies the three requirements that (a) agents observe individual competitors' actions through an observation network; (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, we introduce the idea of *network opacity* – that agents know who they observe, but not who anybody else observes. Instead, we suppose that agents know only the (common) distribution from which those observees are drawn.

This assumption grants an arbitrarily accurate simulation and may therefore 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 the opacity of the network (so that agents are interested in a sequence of weighted average expectations), 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 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 AR(1) process, the full hierarchy of relevant aggregate expectations will follow an ARMA(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 herding behaviour of agents' actions, 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 firms' price-setting decisions, labour search-and-matching models and asset pricing problems. A particular application to firms' price-setting decisions in a classic dynamic pricing environment is examined in depth in the next chapter.

Appendix 2.A Proof of proposition 1.

The Kalman filter (2.10) 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 (2.3), we have that when preparing for period $t+1$, agent i will construct her prior expectation of her social signal as follows:

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

Common knowledge of rationality then allows agent i to substitute in the Kalman filter for agent $\delta_t(i)$'s expectation:

$$E_t(i)\left[E_t(\delta_t(i))\left[X_t\right]\right] = E_t(i)\left[\begin{array}{l} E_{t-1}(\delta_t(i))\left[X_t\right] \\ +K^p\left\{\boldsymbol{s}_t^p(\delta_t(i)) - E_{t-1}(\delta_t(i))\left[\boldsymbol{s}_t^p(\delta_t(i))\right]\right\} \\ +K^s\left\{\begin{array}{l} g_{t-1}(\delta_{t-1}(\delta_t(i))) \\ -E_{t-1}(\delta_t(i))\left[g_{t-1}(\delta_{t-1}(\delta_t(i)))\right] \end{array}\right\} \end{array}\right]$$

The final term shows if agent i is going to observe the period-t action of agent $\delta_t(i)$, then in order to form her prior, she must also consider whomever agent $\delta_t(i)$ observed from period-$(t-1)$. This recursion of expectations (and expectations of expectations) across agents and backwards through time leads to an explosion in the dimensionality (this is the explosion of p) and typically prevents closed-form analysis in anything other than trivially small networks.

However, 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:

$$\begin{aligned} E_t(i)\left[g_{t-1}(\delta_{t-1}(\delta_t(i)))\right] &= \int_0^1 E_t(i)\left[g_{t-1}(j)\right]\phi(j)\,dj \\ &= E_t(i)\left[\int_0^1 g_{t-1}(j)\phi(j)\,dj\right] \\ &= E_t(i)\left[\widetilde{g}_{t-1}\right] \end{aligned}$$

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. Note, too, that by identical logic we also have that when considering their observee's observee, agent i will expect that:

$$E_t(i)\left[E_{t-1}(\delta_t(i))\left[g_{t-1}(\delta_{t-1}(\delta_t(i)))\right]\right] = E_t(i)\left[E_{t-1}(\delta_t(i))\left[\widetilde{g}_{t-2}\right]\right]$$

That is, common knowledge of rationality and the symmetry of agents' problems leads agent i to expect that agent $\delta_t(i)$ makes the same assumption about their own observee. The ongoing recursion backwards through time should be clear. Substituting this all back in above gives:

$$E_t(i)\left[E_t(\delta_t(i))\left[X_t\right]\right] = E_t(i)\begin{bmatrix} E_{t-1}(\delta_t(i))\left[X_t\right] \\ +K^p\left\{\boldsymbol{s}_t^p(\delta_t(i)) - E_{t-1}(\delta_t(i))\left[\boldsymbol{s}_t^p(\delta_t(i))\right]\right\} \\ +K^s\left\{\widetilde{g}_{t-1} - E_{t-1}(\delta_t(i))\left[\widetilde{g}_{t-1}\right]\right\} \end{bmatrix}$$

In effect, this is agent i treating agent $\delta_t(i)$ *as though* (a) they receive a weighted average of everybody's period-$(t-1)$ action and (b) they act in the same manner towards their own observee(s).

So long as the weights used (the observation p.d.f.) are common across agents and constant over time – that is, so long as agents do not learn about the topology of the network – then we have that agent i's problem may be summarised as follows: observe the action of agent $\delta_t(i)$, but treat them as though they, and and all information obtained through them, come from a setting in which all agents observe the weighted average action.

Appendix 2.B Proof of proposition 2.

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

where the weighted sums are defined as:

$$
\begin{aligned}
{}^{\{1\}}\widetilde{\boldsymbol{v}}_{N,t} &\equiv \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{v}_t\left(\delta_t(i)\right) & {}^{\{1\}}\ddot{\boldsymbol{v}}_{N,t} &\equiv \sum_{i=1}^{N} \boldsymbol{v}_t(i)\,\phi_N(i) \\
{}^{\{2\}}\widetilde{\boldsymbol{v}}_{N,t} &\equiv \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{v}_t\left(\delta_t\left(\delta_t(i)\right)\right) & {}^{\{2\}}\ddot{\boldsymbol{v}}_{N,t} &\equiv \sum_{i=1}^{N} \boldsymbol{v}_t\left(\delta_t(i)\right)\phi_N(i) \\
{}^{\{3\}}\widetilde{\boldsymbol{v}}_{N,t} &\equiv \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(\delta_t(i)\right)\right)\right) & {}^{\{3\}}\ddot{\boldsymbol{v}}_{N,t} &\equiv \sum_{i=1}^{N} \boldsymbol{v}_t\left(\delta_t\left(\delta_t(i)\right)\right)\phi_N(i) \\
&\vdots & &\vdots
\end{aligned}
$$

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

Since it is clear that ${}^{\{q\}}\widetilde{\boldsymbol{v}}_{N,t}$ 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.

- ${}^{\{1\}}\widetilde{\boldsymbol{v}}_{N,t} \xrightarrow{d} {}^{\{1\}}\widetilde{\boldsymbol{v}}_t$

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

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

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

$$\begin{aligned} E\left[\boldsymbol{v}_t\left(\delta_t\left(i\right)\right) \boldsymbol{v}_t\left(\delta_t\left(j\right)\right)\right] &= \sum_{k=1}^{N} \phi_N\left(k\right) E\left[\boldsymbol{v}_t\left(k\right) \boldsymbol{v}_t\left(\delta_t\left(j\right)\right)\right] \\ &= \sum_{k=1}^{N} \phi_N\left(k\right) \left(\sum_{l=1}^{N} \phi_N\left(l\right) E\left[\boldsymbol{v}_t\left(k\right) \boldsymbol{v}_t\left(l\right)\right] \right) \\ &= \sum_{k=1}^{N} \phi_N\left(k\right)^2 E\left[\boldsymbol{v}_t\left(k\right) \boldsymbol{v}_t\left(k\right)\right] \\ &= \zeta\left(N\right) \Sigma_{vv} \end{aligned} \tag{2.18}$$

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

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

and thus, in the limit, it must be that

$$\Sigma_{\widetilde{v}\widetilde{v}}^{\{1\}} \equiv \lim_{N\to\infty} Var\left[{}^{\{1\}}\widetilde{\boldsymbol{v}}_{N,t}\right] = \zeta^* \Sigma_{vv} \tag{2.19}$$

- ${}^{\{2\}}\widetilde{\boldsymbol{v}}_{N,t} \xrightarrow{d} {}^{\{2\}}\widetilde{\boldsymbol{v}}_{t}$

The variance of ${}^{\{2\}}\widetilde{\boldsymbol{v}}_{N,t}$ is given by:

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

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 (2.18) 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

$$\sum_{k=1}^{N}\sum_{l\neq k}^{N}\phi_N(k)\,\phi_N(l)=1-\sum_{k=1}^{N}\phi_N(k)^2=1-\zeta(N) \tag{2.20}$$

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

$$E\left[\boldsymbol{v}_t(\delta_t(\delta_t(i)))\,\boldsymbol{v}_t(\delta_t(\delta_t(j)))\right]=\zeta(N)\,\Sigma_{vv}+(1-\zeta(N))\,\zeta(N)\,\Sigma_{vv} \tag{2.21}$$

Substituting this back in, we arrive at

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

and thus, in the limit, it must be that

$$\Sigma_{\widetilde{v}\widetilde{v}}^{\{2\}}\equiv\lim_{N\to\infty}Var\left[{}^{\{2\}}\widetilde{\boldsymbol{v}}_{N,t}\right]=\zeta^*\Sigma_{vv}+(1-\zeta^*)\,\zeta^*\Sigma_{vv} \tag{2.22}$$

- ${}^{\{3\}}\widetilde{\boldsymbol{v}}_{N,t}\xrightarrow{d}{}^{\{3\}}\widetilde{\boldsymbol{v}}_t$

The variance of ${}^{\{3\}}\widetilde{\boldsymbol{v}}_{N,t}$ is given by:

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

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

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

It was shown above in equation (2.21) 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 (2.20), this then implies that when $i \neq j$,

$$\begin{aligned} &E\left[\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(\delta_t\left(i\right)\right)\right)\right)\boldsymbol{v}_t\left(\delta_t\left(\delta_t\left(\delta_t\left(j\right)\right)\right)\right)\right] \\ &\quad = \zeta\left(N\right)\Sigma_{vv} + \left(1-\zeta\left(N\right)\right)\left(\zeta\left(N\right)\Sigma_{vv} + \left(1-\zeta\left(N\right)\right)\zeta\left(N\right)\Sigma_{vv}\right) \end{aligned} \tag{2.23}$$

Substituting this back in, we arrive at

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

and thus, in the limit, it must be that

$$\Sigma_{\widetilde{v}\widetilde{v}}^{\{3\}} \equiv \lim_{N\to\infty} Var\left[{}^{\{3\}}\widetilde{\boldsymbol{v}}_{N,t}\right] = \zeta^*\Sigma_{vv} + \left(1-\zeta^*\right)\left(\zeta^*\Sigma_{vv} + \left(1-\zeta^*\right)\zeta^*\Sigma_{vv}\right) \tag{2.24}$$

- 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:

$$\Sigma_{\widetilde{v}\widetilde{v}}^{\{q\}} = \zeta^*\Sigma_{vv} + \left(1-\zeta^*\right)\Sigma_{\widetilde{v}\widetilde{v}}^{\{q-1\}}$$

This may be simplified by first expanding it as

$$\begin{aligned} \Sigma_{\widetilde{v}\widetilde{v}}^{\{q\}} &= \left(\sum_{p=0}^{q-1}\left(1-\zeta^*\right)^p\right)\zeta^*\Sigma_{vv} \\ &= \left(\frac{1-\left(1-\zeta^*\right)^q}{1-\left(1-\zeta^*\right)}\right)\zeta^*\Sigma_{vv} \\ &= \left(1-\left(1-\zeta^*\right)^q\right)\Sigma_{vv} \end{aligned} \tag{2.25}$$

which completes the proof of the first result.

2. $^{\{q\}}\ddot{\boldsymbol{v}}_{N,t} \xrightarrow{\mathcal{L}^2} {}^{\{q\}}\widetilde{\boldsymbol{v}}_t\ \forall q$

We next demonstrate that $^{\{q\}}\ddot{\boldsymbol{v}}_{N,t}$ converges to $^{\{q\}}\widetilde{\boldsymbol{v}}_t$ in mean square error.[26] That is, we show that $\lim_{N\to\infty} E\left[\left({}^{\{q\}}\ddot{\boldsymbol{v}}_{N,t} - {}^{\{q\}}\widetilde{\boldsymbol{v}}_t\right)^2\right] = 0$. First, see that:

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

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

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

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

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

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

Noting the recursive structure and making use of equation (2.20) then gives us

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

which, in the limit, becomes

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

which is the same rule for $Var\left[{}^{\{q\}}\widetilde{\boldsymbol{v}}_{N,t}\right]$, which implies that

$$\lim_{N\to\infty} Var\left[{}^{\{q\}}\ddot{\boldsymbol{v}}_{N,t}\right] = \lim_{N\to\infty} Var\left[{}^{\{3\}}\widetilde{\boldsymbol{v}}_{N,t}\right] = \Sigma_{\widetilde{v}\widetilde{v}}^{\{q\}}$$

Turning next to the covariance between ${}^{\{q\}}\ddot{\boldsymbol{v}}_{N,t}$ and ${}^{\{q\}}\widetilde{\boldsymbol{v}}_{t}$, we note that

$$\begin{aligned}
Cov\left[{}^{\{q\}}\ddot{\boldsymbol{v}}_{N,t}, {}^{\{q\}}\widetilde{\boldsymbol{v}}_{N,t}\right] &= E\left[\begin{array}{c}\left(\sum_{i=1}^{N}\phi_N\left(i\right)\boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t\left(i\right)))}_{q-1}\right)\right)\\ \times\left(\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t\left(j\right)))}_{q}\right)\right)\end{array}\right]\\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_N\left(i\right)E\left[\boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t\left(i\right)))}_{q-1}\right)\boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t\left(j\right)))}_{q}\right)\right]\\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\phi_N\left(i\right)\phi_N\left(k\right)E\left[\begin{array}{c}\boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t\left(i\right)))}_{q-1}\right)\\ \times\boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t\left(k\right)))}_{q-1}\right)\end{array}\right]
\end{aligned}$$

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

$$Cov\left[{}^{\{q\}}\ddot{\boldsymbol{v}}_{N,t}, {}^{\{q\}}\widetilde{\boldsymbol{v}}_{N,t}\right] = \frac{N}{N}\left(\begin{array}{c}\sum_{i=1}^{N}\phi_N\left(i\right)^2\Sigma_{vv}\\ +\sum_{i=1}^{N}\sum_{k\neq i}^{N}\phi_N\left(i\right)\phi_N\left(k\right)E\left[\begin{array}{c}\boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t\left(i\right)))}_{q-1}\right)\\ \times\boldsymbol{v}_t\left(\underbrace{\delta_t(\cdots(\delta_t\left(k\right)))}_{q-1}\right)\end{array}\right]\end{array}\right)$$

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

$$\lim_{N\to\infty} Cov\left[{}^{\{q\}}\ddot{\boldsymbol{v}}_{N,t}, {}^{\{q\}}\widetilde{\boldsymbol{v}}_{N,t}\right] = \Sigma^{\{q\}}_{\widetilde{v}\widetilde{v}}$$

and, hence, that

$$\begin{aligned}\lim_{N\to\infty} E\left[\left({}^{\{q\}}\ddot{\boldsymbol{v}}_{N,t} - {}^{\{2\}}\widetilde{\boldsymbol{v}}_t\right)^2\right] &= \Sigma^{\{q\}}_{\widetilde{v}\widetilde{v}} - 2\Sigma^{\{q\}}_{\widetilde{v}\widetilde{v}} + \Sigma^{\{q\}}_{\widetilde{v}\widetilde{v}} \\ &= 0\end{aligned}$$

as required.

3. $Cov\left[{}^{\{p\}}\widetilde{\boldsymbol{v}}_t, {}^{\{q\}}\widetilde{\boldsymbol{v}}_t\right] = \Sigma^{\{p\}}_{\widetilde{v}\widetilde{v}}$ $\forall p < q$

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

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

Focussing on the final term, note that

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

Substituting this back into the above then gives

$$\begin{aligned} Cov\left[{}^{\{p\}}\widetilde{\boldsymbol{v}}_t, {}^{\{p+1\}}\widetilde{\boldsymbol{v}}_t\right] &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\phi_N(i)\,\Sigma_{vv} + (1-\phi_N(i))\,\Sigma^p_{\widetilde{v}\widetilde{v}}(N)\right) \\ &= \frac{1}{N}\sum_{i=1}^{N}\left(\phi_N(i)\,\Sigma_{vv} + (1-\phi_N(i))\,\Sigma^p_{\widetilde{v}\widetilde{v}}(N)\right) \\ &= \frac{1}{N}\Sigma_{vv} + \frac{1}{N}\sum_{i=1}^{N}(1-\phi_N(i))\,\Sigma^p_{\widetilde{v}\widetilde{v}}(N) \end{aligned}$$

In the limit, this becomes

$$\lim_{N\to\infty} Cov\left[{}^{\{p\}}\widetilde{\boldsymbol{v}}_{N,t}, {}^{\{p+1\}}\widetilde{\boldsymbol{v}}_{N,t}\right] = \Sigma^p_{\widetilde{v}\widetilde{v}}$$

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

Appendix 2.C Proof of theorem 1.

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

$$X_t \equiv \begin{bmatrix} \boldsymbol{x}_t \\ \overline{E}_t\left[X_t\right] \\ {}^{\{1\}}\widetilde{E}_t\left[X_t\right] \\ {}^{\{2\}}\widetilde{E}_t\left[X_t\right] \\ \vdots \end{bmatrix} = FX_{t-1} + G_1\boldsymbol{u}_t + G_2\boldsymbol{z}_t + G_3\boldsymbol{e}_t + G_4\boldsymbol{z}_{t-1} \tag{2.26}$$

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

$$\boldsymbol{s}_t^p\left(i\right) = D_1\boldsymbol{x}_t + D_2X_{t-1} + R_1\boldsymbol{v}_t\left(i\right) + R_2\boldsymbol{e}_t + R_3\boldsymbol{z}_{t-1} \tag{2.27a}$$

$$\boldsymbol{s}_t^s\left(i\right) = \boldsymbol{\lambda}_1' E_{t-1}\left(\delta_{t-1}\left(i\right)\right)\left[X_{t-1}\right] + \boldsymbol{\lambda}_2'\boldsymbol{x}_{t-1} + \boldsymbol{\lambda}_3'\boldsymbol{v}_{t-1}\left(\delta_{t-1}\left(i\right)\right) \tag{2.27b}$$

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_q} as the matrices that select $\boldsymbol{x}_t$, $\overline{E}_t\left[X_t\right]$ and ${}^{\{q\}}\widetilde{E}_t\left[X_t\right]$ respectively from X_t (e.g., $T_{w_2}X_t = {}^{\{2\}}\widetilde{E}_t\left[X_t\right]$).

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

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

2.C.1 The filter

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

$$\boldsymbol{s}_{t|t-1}(i) \equiv \boldsymbol{s}_t(i) - E_{t-1}(i)[\boldsymbol{s}_t(i)] \tag{2.28}$$

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

$$\begin{aligned} E_t(i)[X_t] &= E[X_t|\mathcal{I}_{t-1}(i)] + E\left[X_t|\boldsymbol{s}_{t|t-1}(i)\right] \\ &= E_{t-1}(i)[X_t] + K_t\boldsymbol{s}_{t|t-1}(i) \end{aligned} \tag{2.29}$$

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, $\boldsymbol{s}_{t|t-1}(i)$, is orthogonal to the projection error, $X_t - K_t\boldsymbol{s}_{t|t-1}(i)$. That is, we require that

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

Rearranging then gives an expression for the optimal Kalman gain:

$$K_t = E\left[X_t\boldsymbol{s}_{t|t-1}(i)'\right]\left(E\left[\boldsymbol{s}_{t|t-1}(i)\,\boldsymbol{s}_{t|t-1}(i)'\right]\right)^{-1} \;\forall i \tag{2.30}$$

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

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

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:

$$\begin{aligned}\boldsymbol{s}_t^p(j) &= (D_1 S_x F + D_2) X_{t-1} + D_1 S_x G_1 \boldsymbol{u}_t \\ &\quad + R_1 \boldsymbol{v}_t(j) + R_2 \boldsymbol{e}_t + R_3 \boldsymbol{z}_{t-1}\end{aligned} \tag{2.31}$$

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

$$E_{t-1}(i)\left[\boldsymbol{s}_t^p(i)\right] = (D_1 S_x F + D_2) E_{t-1}(i)\left[X_{t-1}\right] \tag{2.32}$$

where we have made use of lemma 1 to drop the term in $E_{t-1}(i)[\boldsymbol{z}_{t-1}]$. Subtracting equation (2.32) from (2.31) then gives the innovation in agents' private signals as

$$\begin{aligned}\boldsymbol{s}_{t|t-1}^p(i) &= (D_1 S_x F + D_2) X_{t-1|t-1}(i) + D_1 S_x G_1 \boldsymbol{u}_t \\ &\quad + R_1 \boldsymbol{v}_t(j) + R_2 \boldsymbol{e}_t + R_3 \boldsymbol{z}_{t-1}\end{aligned} \tag{2.33}$$

where $X_{t|t}(i)$ 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

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

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

$$E_{t-1}(i)\left[\boldsymbol{s}_t^s(i)\right] = (\boldsymbol{\lambda}_2' S_x + \boldsymbol{\lambda}_1' T_w) E_{t-1}(i)\left[X_{t-1}\right] \tag{2.34}$$

Subtracting (2.34) from (2.27b), we then have that the innovation in the agent's social signal is given by:

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

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

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

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

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

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

$$\begin{aligned}\boldsymbol{s}_{t|t-1}(i) &= M_1 X_{t-1|t-1}(i) + M_2 X_{t-1|t-1}(\delta_{t-1}(i)) + M_3 X_{t-1} \\ &+ N_1 \boldsymbol{u}_t + N_2 \boldsymbol{v}_t(i) + N_3 \boldsymbol{e}_t + N_4 \boldsymbol{v}_{t-1}(\delta_{t-1}(i)) + N_5 \boldsymbol{z}_{t-1}\end{aligned} \tag{2.35a}$$

where

$$M_1 = \begin{bmatrix} D_1 S_x F + D_2 \\ \boldsymbol{\lambda}_2' S_x + \boldsymbol{\lambda}_1' T_w \end{bmatrix} \quad M_2 = \begin{bmatrix} \mathbf{0} \\ -\boldsymbol{\lambda}_1' \end{bmatrix} \quad M_3 = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\lambda}_1' (I - T_w) \end{bmatrix} \tag{2.35b}$$

$$N_1 = \begin{bmatrix} D_1 S_x G_1 \\ \mathbf{0} \end{bmatrix} \quad N_2 = \begin{bmatrix} R_1 \\ \mathbf{0} \end{bmatrix} \quad N_3 = \begin{bmatrix} R_2 \\ \mathbf{0} \end{bmatrix} \quad N_4 = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\lambda}_3' \end{bmatrix} \quad N_5 = \begin{bmatrix} R_3 \\ \mathbf{0} \end{bmatrix} \tag{2.35c}$$

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

$$\begin{aligned}\boldsymbol{s}_{t|t-1}(i) &= M_1 X_{t-1|t-1}(i) + M_2 \begin{bmatrix} X_{t-1|t-1}(\delta_{t-1}(i,1)) \\ X_{t-1|t-1}(\delta_{t-1}(i,2)) \end{bmatrix} + M_3 X_{t-1} \\ &+ N_1 \boldsymbol{u}_t + N_2 \boldsymbol{v}_t(i) + N_3 \boldsymbol{e}_t + N_4 \begin{bmatrix} \boldsymbol{v}_{t-1}(\delta_{t-1}(i,1)) \\ \boldsymbol{v}_{t-1}(\delta_{t-1}(i,2)) \end{bmatrix} + N_5 \boldsymbol{z}_{t-1}\end{aligned} \tag{2.36a}$$

where

$$M_1 = \begin{bmatrix} D_1 S_x F + D_2 \\ \boldsymbol{\lambda}_2' S_x + \boldsymbol{\lambda}_1' T_w \\ \boldsymbol{\lambda}_2' S_x + \boldsymbol{\lambda}_1' T_w \end{bmatrix} \quad M_2 = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ -\boldsymbol{\lambda}_1' & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{\lambda}_1' \end{bmatrix} \quad M_3 = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\lambda}_1' \left(I - T_w\right) \\ \boldsymbol{\lambda}_1' \left(I - T_w\right) \end{bmatrix} \tag{2.36b}$$

$$N_1 = \begin{bmatrix} D_1 S_x G_1 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \quad N_2 = \begin{bmatrix} R_1 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \quad N_3 = \begin{bmatrix} R_2 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \quad N_4 = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \boldsymbol{\lambda}_3' & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\lambda}_3' \end{bmatrix} \quad N_5 = \begin{bmatrix} R_3 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \tag{2.36c}$$

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 (2.30) as

$$\begin{aligned} E\left[X_t \boldsymbol{s}_{t|t-1}(i)'\right] &= E\begin{bmatrix} \left(F X_{t-1} + G_1 \boldsymbol{u}_t + G_2 \boldsymbol{z}_t + G_4 \boldsymbol{z}_{t-1} + G_3 \boldsymbol{e}_t\right) \\ \times \begin{pmatrix} M_1 X_{t-1|t-1}(i) \\ +M_2 X_{t-1|t-1}\left(\delta_{t-1}(i)\right) \\ +M_3 X_{t-1} \\ +N_1 \boldsymbol{u}_t + N_2 \boldsymbol{v}_t(i) + N_3 \boldsymbol{e}_t \\ +N_4 \boldsymbol{v}_{t-1}\left(\delta_{t-1}(i)\right) + N_5 \boldsymbol{z}_{t-1} \end{pmatrix}' \end{bmatrix} \\ &= E\begin{bmatrix} \left(F X_{t-1}\right)\left(M_1 X_{t-1|t-1}(i)\right)' \\ +\left(F X_{t-1}\right)\left(M_2 X_{t-1|t-1}\left(\delta_{t-1}(i)\right)\right)' \\ +\left(F X_{t-1}\right)\left(M_3 X_{t-1}\right)' \\ +\left(F X_{t-1}\right)\left(N_5 \boldsymbol{z}_{t-1}\right)' \\ +\left(G_1 \boldsymbol{u}_t\right)\left(N_1 \boldsymbol{u}_t\right)' \\ +\left(G_3 \boldsymbol{e}_t\right)\left(N_3 \boldsymbol{e}_t\right)' \\ +\left(G_4 \boldsymbol{z}_{t-1}\right)\left(M_1 X_{t-1|t-1}(i)\right)' \\ +\left(G_4 \boldsymbol{z}_{t-1}\right)\left(M_2 X_{t-1|t-1}\left(\delta_{t-1}(i)\right)\right)' \\ +\left(G_4 \boldsymbol{z}_{t-1}\right)\left(M_3 X_{t-1}\right)' \\ +\left(G_4 \boldsymbol{z}_{t-1}\right)\left(N_5 \boldsymbol{z}_{t-1}\right)' \end{bmatrix} \end{aligned} \tag{2.37}$$

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

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

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 $\left\{\boldsymbol{s}_{t|t-1}(j),\boldsymbol{s}_{t-1|t-2}(j),\cdots\right\}$ and these are orthogonal to $X_{t|t}(j)$ by construction, then it must be that $E_t(j)[X_t]$ and $X_{t|t}(j)$ are orthogonal for all j and t. Note that $V_{t|t}\equiv E\left[X_{t|t}(j)\,X_{t|t}(j)'\right]\ \forall j$ is the variance of each agent's contemporaneous error (common to all agents as their problems are symmetric).

Using this, we may rewrite (2.37) as

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

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

$$\begin{aligned} E\left[X_t \boldsymbol{s}_{t|t-1}(i)'\right] = F & \begin{bmatrix} V_{t-1|t-1} & V_{t-1|t-1} & U_{t-1} \end{bmatrix} M' \\ & + G_1 \Sigma_{uu} N_1' \\ & + F G_2 \Sigma_{zz} N_5' \\ & + G_3 \Sigma_{ee} N_3' \\ & + G_4 \Sigma_{zz} G_2' (M_1 + M_2 + M_3)' \\ & + G_4 \Sigma_{zz} N_5' \end{aligned} \tag{2.38}$$

Turning to the second term in equation (2.30), we have that

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

Expanding out the various cross-products then gives us

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

where $W_{t|t} \equiv E\left[X_{t|t}(i)\,X_{t|t}(j)'\right]$ $\forall i \neq j$ 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

$$\begin{aligned}
E\left[\boldsymbol{s}_{t|t-1}(i)\,\boldsymbol{s}_{t|t-1}(i)'\right] &= M \begin{bmatrix} V_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \\ W_{t-1|t-1} & V_{t-1|t-1} & V_{t-1|t-1} \\ V_{t-1|t-1} & V_{t-1|t-1} & U_{t-1} \end{bmatrix} M' \\
&- M_2 K_{t-1} N_2 \Sigma_{vv} N_4' \\
&- N_4 \Sigma_{vv} N_2' K_{t-1}' M_2' \\
&+ N_1 \Sigma_{uu} N_1' + N_2 \Sigma_{vv} N_2' + N_4 \Sigma_{vv} N_4' \\
&+ (M_1 + M_2 + M_3)\, G_2 \Sigma_{zz} N_5' \\
&+ N_5 \Sigma_{zz} G_2' (M_1 + M_2 + M_3)' \\
&+ N_3 \Sigma_{ee} N_3' \qquad (2.39)
\end{aligned}$$

Substituting (2.38) and (2.39) into (2.30) and gathering like terms, we arrive at:

$$K_t = \begin{pmatrix} F \begin{bmatrix} V_{t-1|t-1} & V_{t-1|t-1} & U_{t-1} \end{bmatrix} M' \\ +G_1 \Sigma_{uu} N_1' \\ +F G_2 \Sigma_{zz} N_5' \\ +G_4 \Sigma_{zz} G_2' \left(M_1 + M_2 + M_3\right)' \\ +G_4 \Sigma_{zz} N_5' \\ +G_3 \Sigma_{ee} N_3' \end{pmatrix}$$
$$\times \begin{bmatrix} M \begin{bmatrix} V_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \\ W_{t-1|t-1} & V_{t-1|t-1} & V_{t-1|t-1} \\ V_{t-1|t-1} & V_{t-1|t-1} & U_{t-1} \end{bmatrix} M' \\ + \left(M_1 + M_2 + M_3\right) G_2 \Sigma_{zz} N_5' \\ +N_5 \Sigma_{zz} G_2' \left(M_1 + M_2 + M_3\right)' \\ -M_2 K_{t-1} N_2 \Sigma_{vv} N_4' \\ -N_4 \Sigma_{vv} N_2' K_{t-1}' M_2' \\ +N_1 \Sigma_{uu} N_1' + N_2 \Sigma_{vv} N_2' + N_4 \Sigma_{vv} N_4' + N_3 \Sigma_{ee} N_3' \end{bmatrix}^{-1} \tag{2.40}$$

2.C.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:

$$\begin{aligned} U_t = {} & F U_{t-1} F' \\ & + G_1 \Sigma_{uu} G_1' + G_2 \Sigma_{zz} G_2' + G_3 \Sigma_{ee} G_3' + G_4 \Sigma_{zz} G_4' + F G_2 \Sigma_{zz} G_4' + G_4 \Sigma_{zz} G_2' F' \end{aligned} \tag{2.41}$$

Variance of agents' expectation errors

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

$$\begin{aligned} X_t - E_{t-1}(i)[X_t] = {} & F \left(X_{t-1} - E_{t-1}(i)[X_{t-1}]\right) \\ & + G_1 \boldsymbol{u}_t + G_2 \boldsymbol{z}_t + G_3 \boldsymbol{e}_t + G_4 \boldsymbol{z}_{t-1} \end{aligned} \tag{2.42}$$

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

$$\begin{aligned} V_{t|t-1} &= FV_{t-1|t-1}F' \\ &\quad + G_1\Sigma_{uu}G_1' + G_2\Sigma_{zz}G_2' + G_3\Sigma_{ee}G_3' + G_4\Sigma_{zz}G_4' + FG_2\Sigma_{zz}G_4' + G_4\Sigma_{zz}G_2'F' \end{aligned} \tag{2.43}$$

Next, we subtract each side of equation (2.29) from X_t and rearrange to obtain

$$\left(X_t - E_t\left(i\right)\left[X_t\right]\right) + K_t\boldsymbol{s}_{t|t-1}\left(i\right) = \left(X_t - E_{t-1}\left(i\right)\left[X_t\right]\right) \tag{2.44}$$

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

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

or

$$V_{t|t} = V_{t|t-1} - K_t \left(\begin{array}{l} \quad M\begin{bmatrix} V_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \\ W_{t-1|t-1} & V_{t-1|t-1} & V_{t-1|t-1} \\ V_{t-1|t-1} & V_{t-1|t-1} & U_{t-1} \end{bmatrix} M' \\ +\left(M_1 + M_2 + M_3\right)G_2\Sigma_{zz}N_5' \\ +N_5\Sigma_{zz}G_2'\left(M_1 + M_2 + M_3\right)' \\ -M_2K_{t-1}N_2\Sigma_{vv}N_4' \\ -N_4\Sigma_{vv}N_2'K_{t-1}'M_2' \\ +N_1\Sigma_{uu}N_1' + N_2\Sigma_{vv}N_2' + N_3\Sigma_{ee}N_3' + N_4\Sigma_{vv}N_4' \end{array}\right) K_t' \tag{2.45}$$

Covariance between agents' expectation errors

First, from (2.42), we have that the prior covariance between two agents' errors is given by:

$$\begin{aligned} W_{t|t-1} &\equiv E\left[X_{t|t-1}\left(i\right)X_{t|t-1}\left(j\right)'\right] \\ &= FW_{t-1|t-1}F' \\ &\quad + G_1\Sigma_{uu}G_1' + G_2\Sigma_{zz}G_2' + G_3\Sigma_{ee}G_3' + G_4\Sigma_{zz}G_4' + FG_2\Sigma_{zz}G_4' + G_4\Sigma_{zz}G_2'F' \end{aligned} \tag{2.46}$$

Next, returning to equation (2.44)

$$(X_t - E_t(i)[X_t]) = (X_t - E_{t-1}(i)[X_t]) - K_t \boldsymbol{s}_{t|t-1}(i) \tag{2.47}$$

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

$$\begin{aligned} W_{t|t} = W_{t|t-1} &\\ + K_t \, Cov\left(\boldsymbol{s}_{t|t-1}(i), \boldsymbol{s}_{t|t-1}(j)\right) K_t' &\\ - Cov\left(X_{t|t-1}(i), \boldsymbol{s}_{t|t-1}(j)\right) K_t' &\\ - K_t \, Cov\left(\boldsymbol{s}_{t|t-1}(i), X_{t|t-1}(j)\right) & \end{aligned} \tag{2.48}$$

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

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

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

$$\begin{aligned}
E\left[\boldsymbol{s}_{t|t-1}(i)\,\boldsymbol{s}_{t|t-1}(j)'\right] = M &\begin{bmatrix} W_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \\ W_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \\ V_{t-1|t-1} & V_{t-1|t-1} & U_{t-1} \end{bmatrix} M' \\
&+ (M_1 + M_2 + M_3)\, G_2 \Sigma_{zz} N_5' \\
&+ N_5 \Sigma_{zz} G_2' (M_1 + M_2 + M_3)' \\
&+ N_1 \Sigma_{uu} N_1' \\
&+ N_3 \Sigma_{ee} N_3' \qquad (2.49)
\end{aligned}$$

For the third term, we have

$$\begin{aligned}
Cov\left(X_{t|t-1}(i), \boldsymbol{s}_{t|t-1}(j)\right) &= E\left[\begin{array}{c} \begin{pmatrix} FX_{t-1|t-1}(j) \\ +G_1\boldsymbol{u}_t \\ +G_2\boldsymbol{z}_t \\ +G_4\boldsymbol{z}_{t-1} \\ +G_3\boldsymbol{e}_t \end{pmatrix} \\ \times \begin{pmatrix} M_1 X_{t-1|t-1}(i) \\ +M_2 X_{t-1|t-1}(\delta_{t-1}(i)) \\ +M_3 X_{t-1} \\ +N_1\boldsymbol{u}_t + N_2\boldsymbol{v}_t(i) \\ +N_4\boldsymbol{v}_{t-1}(\delta_{t-1}(i)) + N_5\boldsymbol{z}_{t-1} + N_3\boldsymbol{e}_t \end{pmatrix}' \end{array}\right] \\
&= F\begin{bmatrix} V_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \end{bmatrix} M' \\
&+ G_1 \Sigma_{uu} N_1' \\
&+ FG_2 \Sigma_{zz} N_5' \\
&+ G_4 \Sigma_{zz} G_2' (M_1 + M_2 + M_3)' \\
&+ G_4 \Sigma_{zz} N_5' \\
&+ G_3 \Sigma_{zz} N_3' \qquad (2.50)
\end{aligned}$$

while the fourth term is the transpose of the same.

Filter summary

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

$$\begin{aligned}
E\left[\boldsymbol{s}_{t|t-1}(i)\,\boldsymbol{s}_{t|t-1}(i)'\right] = M &\begin{bmatrix} V_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \\ W_{t-1|t-1} & V_{t-1|t-1} & V_{t-1|t-1} \\ V_{t-1|t-1} & V_{t-1|t-1} & U_{t-1} \end{bmatrix} M' \\
&+ (M_1 + M_2 + M_3)\, G_2 \Sigma_{zz} N_5' \\
&+ N_5 \Sigma_{zz} G_2' (M_1 + M_2 + M_3)' \\
&- M_2 K_{t-1} N_2 \Sigma_{vv} N_4' \\
&- N_4 \Sigma_{vv} N_2' K_{t-1}' M_2' \\
&+ N_1 \Sigma_{uu} N_1' + N_2 \Sigma_{vv} N_2' + N_4 \Sigma_{vv} N_4'
\end{aligned} \tag{2.51a}$$

$$\begin{aligned}
E\left[\boldsymbol{s}_{t|t-1}(i)\,\boldsymbol{s}_{t|t-1}(j)'\right] = M &\begin{bmatrix} W_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \\ W_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \\ V_{t-1|t-1} & V_{t-1|t-1} & U_{t-1} \end{bmatrix} M' \\
&+ (M_1 + M_2 + M_3)\, G_2 \Sigma_{zz} N_5' \\
&+ N_5 \Sigma_{zz} G_2' (M_1 + M_2 + M_3)' \\
&+ N_1 \Sigma_{uu} N_1'
\end{aligned} \tag{2.51b}$$

$$\begin{aligned}
E\left[X_t \boldsymbol{s}_{t|t-1}(i)'\right] = F &\begin{bmatrix} V_{t-1|t-1} & V_{t-1|t-1} & U_{t-1} \end{bmatrix} M' \\
&+ G_1 \Sigma_{uu} N_1' \\
&+ F G_2 \Sigma_{zz} N_5' \\
&+ G_4 \Sigma_{zz} G_2' (M_1 + M_2 + M_3)' \\
&+ G_4 \Sigma_{zz} N_5'
\end{aligned} \tag{2.51c}$$

$$\begin{aligned}
E\left[X_{t|t-1}(i)\,\boldsymbol{s}_{t|t-1}(j)'\right] = F &\begin{bmatrix} V_{t-1|t-1} & W_{t-1|t-1} & V_{t-1|t-1} \end{bmatrix} M' \\
&+ G_1 \Sigma_{uu} N_1' \\
&+ F G_2 \Sigma_{zz} N_5' \\
&+ G_4 \Sigma_{zz} G_2' (M_1 + M_2 + M_3)' \\
&+ G_4 \Sigma_{zz} N_5'
\end{aligned} \tag{2.51d}$$

$$K_t = E\left[X_t \boldsymbol{s}_{t|t-1}(i)'\right]\left(E\left[\boldsymbol{s}_{t|t-1}(i)\,\boldsymbol{s}_{t|t-1}(i)'\right]\right)^{-1} \tag{2.51e}$$

$$\begin{aligned} U_t &= FU_{t-1}F' \\ &\quad + G_1\Sigma_{uu}G_1' + G_2\Sigma_{zz}G_2' + G_4\Sigma_{zz}G_4' + FG_2\Sigma_{zz}G_4' + G_4\Sigma_{zz}G_2'F' \end{aligned} \tag{2.51f}$$

$$\begin{aligned} V_{t|t-1} &= FV_{t-1|t-1}F' \\ &\quad + G_1\Sigma_{uu}G_1' + G_2\Sigma_{zz}G_2' + G_4\Sigma_{zz}G_4' + FG_2\Sigma_{zz}G_4' + G_4\Sigma_{zz}G_2'F' \end{aligned} \tag{2.51g}$$

$$\begin{aligned} W_{t|t-1} &= FW_{t-1|t-1}F' \\ &\quad + G_1\Sigma_{uu}G_1' + G_2\Sigma_{zz}G_2' + G_4\Sigma_{zz}G_4' + FG_2\Sigma_{zz}G_4' + G_4\Sigma_{zz}G_2'F' \end{aligned} \tag{2.51h}$$

$$V_{t|t} = V_{t|t-1} - K_t\, E\left[\boldsymbol{s}_{t|t-1}(i)\,\boldsymbol{s}_{t|t-1}(i)'\right] K_t' \tag{2.51i}$$

$$\begin{aligned} W_{t|t} = W_{t|t-1} &+ K_t E\left[\boldsymbol{s}_{t|t-1}(i)\,\boldsymbol{s}_{t|t-1}(j)'\right] K_t' \\ &- E\left[X_{t|t-1}(i)\,\boldsymbol{s}_{t|t-1}(j)'\right] K_t' \\ &- K_t E\left[\boldsymbol{s}_{t|t-1}(i)\,X_{t|t-1}(j)'\right] \end{aligned} \tag{2.51j}$$

Provided has 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.

2.C.3 Confirming the conjectured law of motion

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

$$X_t \equiv \begin{bmatrix} \boldsymbol{x}_t \\ \overline{E}_t [X_t] \\ {}^{\{1\}}\widetilde{E}_t [X_t] \\ {}^{\{2\}}\widetilde{E}_t [X_t] \\ \vdots \end{bmatrix} = F X_{t-1} + G_1 \boldsymbol{u}_t + G_2 \boldsymbol{z}_t + G_3 \boldsymbol{e}_t + G_4 \boldsymbol{z}_{t-1} \tag{2.52}$$

To confirm this law of motion, we first combining equations (2.29) and (2.36) 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

$$\begin{aligned} E_t (i) [X_t] = {} & K (M_1 + M_2 + M_3) X_{t-1} \\ & + (F - K M_1) E_{t-1} (i) [X_{t-1}] \\ & - K M_2 E_{t-1} (\delta_{t-1} (i)) [X_{t-1}] \\ & + K N_1 \boldsymbol{u}_t \\ & + K N_2 \boldsymbol{v}_t (i) \\ & + K N_3 \boldsymbol{e}_t \\ & + K N_4 \boldsymbol{v}_{t-1} (\delta_{t-1} (i)) \\ & + K N_5 \boldsymbol{z}_{t-1} \end{aligned} \tag{2.53}$$

Taking the simple average of equation (2.53) gives

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

where I have used proposition 2 to replace $\int_0^1 \boldsymbol{v}_{t-1}\left(\delta_{t-1}(i)\right) di$ with ${}^{\{1\}}\widetilde{\boldsymbol{v}}_{t-1}$. But since ${}^{\{1\}}\widetilde{\boldsymbol{v}}_{t-1}$ is part of $\boldsymbol{z}_{t-1}$, while $\overline{E}_{t-1}\left[X_{t-1}\right]$ and ${}^{\{1\}}\widetilde{E}_{t-1}\left[X_{t-1}\right]$ are part of X_{t-1}, we can simplify this down to:

$$\begin{aligned}
\overline{E}_t\left[X_t\right] = & \left\{K\left(M_1+M_2+M_3\right)+\left(F-K M_1\right) T_s-K M_2 T_{w_1}\right\} X_{t-1} \\
& +K N_1 \boldsymbol{u}_t \\
& +K N_3 \boldsymbol{e}_t \\
& +K\left(\begin{bmatrix} N_4 & \mathbf{0}_{1\times\infty} \end{bmatrix}+N_5\right) \boldsymbol{z}_{t-1}
\end{aligned} \tag{2.54}$$

Next, taking the q-th weighted average of equation (2.53) gives

$$\begin{aligned}
{}^{\{q\}}\widetilde{E}_t\left[X_t\right] = & K\left(M_1+M_2+M_3\right) X_{t-1} \\
& +\left(F-K M_1\right){}^{\{q\}} \widetilde{E}_{t-1}\left[X_{t-1}\right] \\
& -K M_2{}^{\{q+1\}} \widetilde{E}_{t-1}\left[X_{t-1}\right] \\
& +K N_1 \boldsymbol{u}_t \\
& +K N_2{}^{\{q\}} \widetilde{\boldsymbol{v}}_t \\
& +K N_3 \boldsymbol{e}_t \\
& +K N_4{}^{\{q+1\}} \widetilde{\boldsymbol{v}}_{t-1} \\
& +K N_5 \boldsymbol{z}_{t-1}
\end{aligned}$$

where the last two terms have again made use of proposition 2. From this, we can read immediately that

$$
\begin{aligned}
{}^{\{q\}}\widetilde{E}_t\left[X_t\right] = &\left\{K\left(M_1+M_2+M_3\right)+\left(F-KM_1\right)T_{w_q}-KM_2T_{w_{q+1}}\right\}X_{t-1}\\
&+KN_1\boldsymbol{u}_t\\
&+K\begin{bmatrix}\mathbf{0}_{1\times r(q-1)} & N_2 & \mathbf{0}_{1\times\infty}\end{bmatrix}\boldsymbol{z}_t\\
&+KN_3\boldsymbol{e}_t\\
&+K\left(\begin{bmatrix}\mathbf{0}_{1\times rq} & N_4 & \mathbf{0}_{1\times\infty}\end{bmatrix}+N_5\right)\boldsymbol{z}_{t-1}
\end{aligned}
\tag{2.55}
$$

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 (2.54) and (2.55) into equation (2.52) to arrive at

$$
F=\begin{bmatrix}
\begin{bmatrix}A & \mathbf{0}_{m\times\infty}\end{bmatrix}\\
K\left(M_1+M_2+M_3\right)+\left(F-KM_1\right)T_s-KM_2T_{w_1}\\
K\left(M_1+M_2+M_3\right)+\left(F-KM_1\right)T_{w_1}-KM_2T_{w_2}\\
K\left(M_1+M_2+M_3\right)+\left(F-KM_1\right)T_{w_2}-KM_2T_{w_3}\\
\vdots
\end{bmatrix}
\tag{2.56a}
$$

$$
G_1=\begin{bmatrix}P\\ KN_1\\ KN_1\\ KN_1\\ \vdots\end{bmatrix}\qquad
G_2=\begin{bmatrix}
\mathbf{0}_{m\times\infty}\\
\mathbf{0}_{\infty\times\infty}\\
K\begin{bmatrix}N_2 & \mathbf{0}_{1\times r} & \mathbf{0}_{1\times r} & \mathbf{0}_{1\times\infty}\end{bmatrix}\\
K\begin{bmatrix}\mathbf{0}_{1\times r} & N_2 & \mathbf{0}_{1\times r} & \mathbf{0}_{1\times\infty}\end{bmatrix}\\
\vdots
\end{bmatrix}
\tag{2.56b}
$$

$$
G_3=\begin{bmatrix}\mathbf{0}_{m\times n}\\ KN_3\\ KN_3\\ KN_3\\ \vdots\end{bmatrix}\qquad
G_4=\begin{bmatrix}
\mathbf{0}_{m\times\infty}\\
K\left(\begin{bmatrix}N_4 & \mathbf{0}_{1\times p} & \mathbf{0}_{1\times r} & \mathbf{0}_{1\times\infty}\end{bmatrix}+N_5\right)\\
K\left(\begin{bmatrix}\mathbf{0}_{1\times p} & N_4 & \mathbf{0}_{1\times r} & \mathbf{0}_{1\times\infty}\end{bmatrix}+N_5\right)\\
K\left(\begin{bmatrix}\mathbf{0}_{1\times r} & \mathbf{0}_{1\times r} & N_4 & \mathbf{0}_{1\times\infty}\end{bmatrix}+N_5\right)\\
\vdots
\end{bmatrix}
\tag{2.56c}
$$

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 (2.56) 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.

Appendix 2.D 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:

$$g_t\left(i\right) = \boldsymbol{\alpha}'\boldsymbol{s}_t^p\left(i\right) + \boldsymbol{\eta}_x' E_t\left(i\right)\left[X_t\right] + \eta_y E_t\left(i\right)\left[\overline{g}_t\right] + \eta_z E_t\left(i\right)\left[\overline{g}_{t+1}\right] \tag{2.57}$$

where agents' private signals are formed as

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

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 $g_t\left(i\right)$ may be expressed in the general form

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

To do this, we start by taking the simple average of equation (2.57) 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:

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

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

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

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 $\boldsymbol{x}_t$):

$$\boldsymbol{x}_t = S X_t$$

Then we can write:

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

or, in general,

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

The average period-t action can therefore be written as

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

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

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

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

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

for some matrix of parameters F. This implies that

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

and hence that

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

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

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

Continued substitution then arrives at:

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

which, in turn, becomes

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

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

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

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

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

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):

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

Chapter 3

Networks and Inflation

Abstract

This paper presents a model of price setting wherein firms partially inform their decisions by watching price changes by other firms across an observation network. Within a context of imperfect common knowledge and for a wide range of plausible and commonly observed network structures, idiosyncratic shocks are shown to not "wash out" in aggregate prices. These aggregate effects are also shown to be persistent despite the underlying idiosyncratic shocks being entirely transitory and firms having complete flexibility in their price-setting. The model is therefore able to explain a variety of recently documented stylised facts regarding price setting, including the observation that short-lived price changes appear to contain macroeconomic content.

3.1 Introduction

This paper develops a network learning-based microfoundation for cost-push shocks, with the aggregate price level able to persistently deviate from it's long-run trend despite (a) the absence of any aggregate shocks to the economy; (b) firms being free to adjust their prices every period; and (c) network shocks (comprised of weighted sums of idiosyncratic shocks) being purely transitory.

That idiosyncratic shocks are an important aspect of firms' price-setting decisions is now universally accepted. However, it remains commonly assumed that since the shocks themselves must cancel out,[1] the *effects* of those shocks on firms' decisions

[1] Over a continuum of agents, that mean-zero idiosyncratic shocks must sum to zero is true by definition; if they did not, they would necessarily include an aggregate component.

must also wash out in aggregation. In such a setting, firm-specific shocks can only contribute to aggregate dynamics by causing sluggish responses to aggregate shocks, because firms take time to be sure that a given shock is truly common to all firms. In contrast to this, recently documented evidence from studies of micro-level price changes suggests that those price changes most likely to have been driven by idiosyncratic factors do *not* cancel out and therefore do indeed appear to contain content of macroeconomic importance.

To achieve the emergence of aggregate effects from idiosyncratic shocks, this chapter makes applied use of the model developed in the previous chapter: we suppose that firms with complete price-setting flexibility learn about the state of the economy by observing each other's prices in a directed network and set their own prices on the basis of their marginal costs and, for reasons of strategic complementarity, their beliefs regarding the average price.

With unobserved aggregate variables following an AR(1) process, we show that the full hierarchy of firms' expectations will follow an ARMA(1,1) process, with current and lagged weighted averages of firms' idiosyncratic shocks entering at an aggregate level. For sufficiently irregular networks (i.e. when the link distribution is sufficiently non-uniform) these weighted sums are shown to not converge to zero, thereby adding aggregate volatility to the system. Despite idiosyncratic shocks being purely transitory, the aggregate volatility they induce through the network is also shown to exhibit (endogenous) persistence.

This chapter therefore adds to the burgeoning literature on deriving aggregate volatility from agents' idiosyncratic shocks. Key in this field to date includes the work by Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Saleh (2012).[2] Examining the idea of firms operating within an inter-sectoral supply network, they demonstrate idiosyncratic productivity shocks leading to volatility in aggregate output and, for finite networks, derive an upper limit for the rate at which aggregate volatility declines as the number of firms increases. For sufficiently asymmetric trading networks, they show that aggregate volatility need not vanish at all. In another vein, Gabaix (2011) demonstrates how aggregate volatility can emerge from idiosyncratic shocks

[2]The work of this paper was first developed independently by Carvalho (2010) and Acemoglu, Ozdaglar, and Tahbaz-Saleh (2010) and later combined to the paper referenced in the text.

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 chapter 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. In the current paper, with network-based learning, it derives from firms whose price changes are most readily observed.

It may be noted that because firms may choose to observe the prices of other firms with whom they do not trade and are not competitors (a perfectly reasonable action provided that their marginal costs are correlated), this model also represents a novel transmission mechanism for inflation across industries or geographies independent of it's path along production chains. However, the origin of an observation network remains largely outside the scope of the current paper, which takes the network as exogenously given.

When firms exist in an observation network that includes (a) repeated actions; (b) Bayesian updating; and (c) strategic interaction, it becomes necessary for them to estimate not only the simple average of all firms' expectations (for reasons of strategic interaction), but also the expectations of their individual observees and, in turn, their observees' expectations of others again. 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 1.2 in the first chapter for more detail) and has typically been thought to prevent closed-form analysis in anything other than trivially small networks.

To date, research in network learning has therefore abandoned one of these three assumptions in order to achieve tractability. For example, in abandoning the assumption of repeated actions, Acemoglu, Dahleh, Lobel, and Ozdaglar (2011) characterise the (Bayesian) equilibrium of a sequential learning model for a general stochastic sampling process and 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. In giving up on rational (Bayesian) learning, Golub and Jackson (2010) 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 in this sub-field, they are able to 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. Finally, Mueller-Frank (2013) details a formal structure for Bayesian learning over an undirected social network (i.e. with pairwise sharing) in the absence of strategic concerns by agents. He notes the extreme practical difficulties of actually implementing such a rule, both for the agents in principle and the researcher more generally.

In contrast to these, this paper makes use of results presented in the previous chapter that permit the inclusion of all three assumptions regarding network learning by combining them with a fourth: network opacity. By denying agents knowledge of the exact topology of the network and instead supposing that they know only the (i.i.d.) distribution from which observation targets are drawn and do not learn about the structure of the network over time, firms' state vector of interest includes an infinite sequence of *weighted average* expectations instead of individual agents' expectations. Because of the recursive nature of firms' learning, this sequence will be of decreasing importance to the hierarchy of simple-average expectations, so an arbitrarily accurate approximation of the full solution may then be found by selecting a sufficiently high cut-off for the number of weighted-average expectations to include, together with the standard cut-off for the number of higher-orders of expectation.

This paper falls broadly within and was initially inspired by the literature on imperfect common knowledge (where firms possess incomplete information about aggregate state variables because of imperfect signals). The idea that real effects may arise from nominal disturbances through imperfect information dates to Lucas (1972) and, more recently, Woodford (2003). The solution method developed by this paper builds upon that put forward by Nimark (2008, 2011a), who introduced dynamic pricing and idiosyncratic shocks in marginal costs to the Woodford (2003) paper. Other recent work in this area includes Adam (2007), who looked at optimal monetary policy in the Woodford setting and Melosi (2012), who uses the Survey of Professonal Forecasters to estimate a DSGE model with price setters experiencing imperfect common knowledge.

Of course, firms existing within observation networks need not only feed into a setting of imperfect common knowledge. It might also, for example, be readily applied to the rational inattention work of Sims (2003) or the "sticky information" literature of Mankiw and Reis (2002) and Reis (2006). In this latter case, if we suppose that a full information update is costly and observing the price of a competitor less so, it is easy to see that a natural incentive emerges for a firm to delay a full update and instead "free ride" on the price changes of their competitors. Exploration of network learning in these other settings would be a fruitful area of research. However, as shown below, evidence from a variety of surveys of firms' price-setting behaviour suggests that the imperfect common knowledge setting may be the more likely reason for firms' observation of each others' prices.

The remainder of this paper is organised as follows. Section 3.2 provides evidence of the key assumptions of this paper. Section 3.3 then presents a DSGE model in which firms are free to adjust their prices every period, but suffer from incomplete information and seek to remedy this by observing each others' prices in a network. Section 3.4 presents simulation results, while section 3.5 concludes.

3.2 Evidence

In this section we gather evidence in support of the key assumptions of this chapter's model. In particular, we argue here that (a) firms set their prices, in part, on the basis of observed prices posted by individual competitors; and (b) this is done within a context of imperfect common knowledge, in the sense of Woodford (2003).

That firms operate within not just transactional but also *observational* networks is to some extent intuitive, or even self evident. An independent coffee shop will obviously take note of the prices offered by their competitors, including both other independent outlets nearby and larger chains such as Starbucks. The model presented here goes further than this, however, because firms might also observe the price movements of businesses that are not direct competitors or suppliers in order to learn about factors common to all firms. When the manager of a book shop observes a price change at a Thai restaurant next door, for example, or even a car mechanic around the corner, they obtain information about movements in average marginal costs and local demand, thereby improving their ability to ascertain that portion of their own cost or demand changes that are idiosyncratic.

We here first describe evidence from a number of price-setting surveys conducted (typically by or on behalf of central banks) in the 1990s and 2000s and then explore what evidence may be garnered from recent studies of directly observed price changes.

3.2.1 Price-setting surveys

Starting with the work of Blinder (1991) and Blinder, Canetti, Lebow, and Rudd (1998) in the United States and continuing through to the first half of the 2000s, a variety of surveys were conducted in an attempt to shed light on precisely how firms set prices. These include work in the UK (Hall, Walsh, and Yates, 1997); Sweden (Apel, Friberg, and Hallsten, 2005); Japan (Nakagawa, Hattori, and Takagawa, 2000); Canada (Amirault, Kwan, and Wilkinson, 2006); and nine euro-zone countries (Fabiani, Druant, Hernando, Kwapil, Landau, Loupias, Martins, Mathä, Sabbatini, Stahl, and Stokman, 2005).[3]

[3]Countries included were: Austria (Kwapil, Baumgartner, and Scharler, 2005); Belgium (Aucremanne and Druant, 2005); France (Loupias and Ricart, 2004); Germany (Stahl, 2005); Luxembourg

When looking at those firms following partially or completely state-based pricing,[4] Canadian firms listed price changes by competitors as the most important cause in triggering an adjustment, as did those in Sweden. In Spain, 53% of firms reported that competitors' price movements were important factors in triggering their own price changes. In considering the *magnitude* of price changes, 25% of surveyed UK firms reported basing their prices on those of their competitors. This figure agreed with the 27% of surveyed eurozone firms reporting the same, although this ranged from 13% in Portugal to 38% in France. In the Netherlands, where the survey was unique in including very small firms among those polled, this figure was 21.6% overall but rose sharply to 34.1% for firms employing only one worker.

These responses are strongly supportive of the idea that firms observe each others' prices, and we can only assume that they do so as a result of some form of imperfect information; that they learn something from their observations.

However, that firms observe each others' prices does not, in itself, speak to *why* they might do so. If, for example, firms experience significant costs in gathering information and developing optimal price plans in the style of the "sticky information" models of Mankiw and Reis (2002, 2006, 2007) and Reis (2006), then mimicking the price changes of one's competitors may be a useful short-cut. A fair approximation of the firm's optimal price could then be achieved by observing the prices of competitors with similar production technologies and who face similar demand. Alternatively, if firms face strategic complementarity in their price-setting and there are unobservable aggregate state variables in the style of Woodford (2003) and Nimark (2008), observing other firms' decisions may be used to inform businesses of the average actions or beliefs of their competitors.

Fortunately, the surveys also queried firms as to their opinions regarding the reasons for price stickiness, from which four theories stand out as being significant: implicit contracts, explicit contracts, cost-based pricing and coordination failure. All of these were among the top five recognised reasons in all 14 surveys when they were

(Lünnemann and Mathä, 2006); the Netherlands (Hoeberichts and Stokman, 2010); Portugal (Martins, 2005); and Spain (Alvarez and Hernando, 2005).

[4]The alternative being to use entirely time-based pricing, whereby firms adjust their prices at a set (average) frequency irrespective of economic conditions.

included in the options put to surveyed firms. In stark contrast, menu costs and its more recent variant, information costs, were among the least supported ideas, being in the bottom three reasons for most European surveys and Canada. Only in America and Austria were these costs placed in the middle of the group, menu costs being cited as the sixth most proximate cause of price rigidity in the United States and seventh in Austria and information costs coming sixth in Austria.

The low importance attached to information costs suggests that while there may be imperfect information, it does not manifest in the form of infrequently updated information sets. On the contrary, the strong recognition of coordination concerns and cost-based pricing are supportive of this paper's underlying model: the former suggests that businesses are concerned with their strategic complementarity in price-setting and the latter that (presumably marginal) costs drive movements in prices.

3.2.2 Stylised facts from observed price changes

Although early work suggested that most prices change around once per year,[5] the seminal work by Bils and Klenow (2004) observed that the median duration of prices in CPI data from the U.S. Bureau of Labor Statistics (BLS) was 4.3 months, a frequency almost three times higher than previously thought. This triggered a rush of further work exploring and broadly characterising microeconomic price changes. Klenow and Malin (2010) provide an excellent survey of this literature and provide a summary in the form of ten stylised facts. Among these are that:

1. prices change at least once a year, twice in America;
2. temporary price changes – both reductions and increases – around more rigid "reference prices" are common and do not cancel out in aggregation, suggesting that some macroeconomic content is present in the more frequent updates;
3. price changes are typically larger than those needed to keep up with inflation, suggesting that idiosyncratic factors weigh more heavily on a firm's price-setting decision than aggregate factors;
4. changes in *relative* prices tend to be short lived, suggesting that idiosyncratic shocks are less persistent than aggregate disturbances; and

[5] See, for example, Taylor (1999).

5. price changes are generally linked to changes in marginal costs, particularly wages.

The first of these dictates that we require some form of structural, or *real* rigidity in addition to firms' nominal rigidities – a "contract multiplier," in the words of Taylor (1980) – to explain the sluggish responses observed in aggregate price indicies.[6] The second and third points suggest that even if firms' idiosyncratic shocks have zero mean and "cancel out" when averaged, average temporary price changes (that are presumably based upon them) do not cancel out. Finally, the fourth and fifth points are suggestive of a model in which firms' marginal costs are subject to persistent aggregate shocks and only transitory idiosyncratic shocks.

The model presented in this paper is consistent with all of the above stylised facts and with observations of rigidity in aggregate prices. Because firms are able to observe the prices of any other firm, it also represents a framework for the transmission of inflation (and hence, its persistence) across industries or geographies and not simply along production chains.

[6] See, for example, Christiano, Eichenbaum, and Evans (1999) or Romer and Romer (2004) for the USA, or Peersman and Smets (2003) for the Euro area.

3.3 The Model

In this section we construct and analyse a dynamic, stochastic, general equilibrium (DSGE) model in which firms make use of network learning in their pricing decisions. The real economy is presented here in a standard model with no capital. A representative household purchases differentiated goods via a Dixit and Stiglitz (1977) aggregator and supplies labour to firms. Monopolistic firms produce the goods and sell them to the household, with prices able to be adjusted costlessly every period. Persistent aggregate shocks occur within the household's preferences, the central bank's interest rate policy and economy-wide TFP, while firms also face idiosyncratic, transitory shocks to their demand, nominal wages and productivity. Only the main results are presented here; readers interested in the full derivation are referred to appendix 3.A.

In what follows, unless otherwise indicated, lower-case letters are used to denote (natural) log deviations from the long-run steady state values of the corresponding upper-case variables (e.g. $y_t \equiv \ln(Y_t) - \ln(Y^{ss})$).

3.3.1 The household

Each period, a representative household maximises

$$E_t^{HH}\left[\sum_{s=0}^{\infty}\beta^s\left\{e^{\epsilon_{Ct+s}}\frac{C_{t+s}^{1-\frac{1}{\sigma}}-1}{1-\frac{1}{\sigma}}-e^{\epsilon_{Ht+s}}\frac{H_{t+s}^{1+\frac{1}{\psi}}}{1+\frac{1}{\psi}}\right\}\right] \tag{3.1}$$

subject to a standard budget constraint and where $E_t^{HH}[\cdot]$ is the mathematical expectation conditional on the household's information set in period t (defined below); C_t is aggregate consumption; H_t is the aggregate labour supply; σ is the elasticity of intertemporal substitution; ψ is the Frisch elasticity of labour supply; and ϵ_{Ct} and ϵ_{Ht} are persistent, mean zero shocks (specified below) to the utility of consumption and the disutility of labour respectively. The shock to the disutility of labour may be considered a reduced-form way of capturing broad shocks to the labour supply, such as a temporary impairment to labour mobility.

Aggregate consumption is given by the Dixit-Stiglitz aggregator over individual consumption goods:

$$C_t = \left(\int \left(e^{-v_{y,t}(j)} C_t(j) \right)^{\frac{\varepsilon-1}{\varepsilon}} dj \right)^{\frac{\varepsilon}{\varepsilon-1}} \tag{3.2}$$

where ε is the elasticity of substitution and $v_{y,t}(j)$ is a transitory, mean zero, idiosyncratic shock to the household's demand for good j (defined below). The household's subsequent first-order conditions are:

$$\frac{W_t}{P_t} e^{\epsilon_{Ct}} C_t^{-\frac{1}{\sigma}} = e^{\epsilon_{Ht}} H_t^{\frac{1}{\psi}} \tag{3.3}$$

$$e^{\epsilon_{Ct}} C_t^{-\frac{1}{\sigma}} = \beta (1 + i_t) E_t^{HH} \left[e^{\epsilon_{Ct+1}} C_{t+1}^{-\frac{1}{\sigma}} \frac{1}{\Pi_{t+1}} \right] \tag{3.4}$$

where W_t/P_t is the real wage; i_t is the net nominal interest rate; and $\Pi_t \equiv P_t/P_{t-1}$ is the gross rate of inflation. It can also be shown that household demand for good j is given by:

$$C_t(j) = \left(\frac{P_t(j)}{P_t} \right)^{-\varepsilon} C_t e^{v_{y,t}(j)} \tag{3.5}$$

and the aggregate price level by:

$$P_t = \left(\int P_t(j)^{1-\varepsilon} dj \right)^{\frac{1}{1-\varepsilon}} \tag{3.6}$$

3.3.2 Firms

Each good is produced by a single firm according to a common production function that deploys labour with decreasing marginal productivity:

$$Y_t(j) = A_t(j) H_t(j)^{1-\alpha} \tag{3.7}$$

with each firm's productivity, $A_t(j)$, given by:

$$\ln(A_t(j)) = \epsilon_{At} + v_{a,t}(j) \tag{3.8}$$

where ϵ_{At} is a persistent, mean zero, aggregate shock and $v_{a,t}(j)$ is a transitory, mean zero, idiosyncratic shock (each specified below) to the firm's productivity, broadly defined. Firm j's real marginal cost is then:

$$MC_t(j) = (1 + \eta) \frac{W_t(j)}{P_t} \frac{1}{A_t(j)} \left(\frac{Y_t(j)}{A_t(j)} \right)^{\eta} \tag{3.9}$$

where $\eta \equiv \frac{\alpha}{1-\alpha}$ is the elasticity of marginal cost w.r.t. output and $W_t(j)$ is the nominal wage paid by the firm, defined as:

$$W_t(j) \equiv W_t e^{v_{w,t}(j)} \tag{3.10}$$

where $v_{w,t}(j)$ is a transitory, mean zero shock to the firm's wage bargaining. Shocks to $A_t(j)$ should therefore be broadly interpreted as a reduced-form means of capturing shocks to firms' marginal costs other than those that act through demand or the (real) wage.[7]

Firms engage in static pricing – i.e. they are free to costlessly update their prices in every period – so that results presented here represent real rigidity, not nominal. Given the results of Bils and Klenow (2004) and subsequent research, it is arguably best to assume that there exists no nominal rigidity in at least some industries.

In this case, firm j's optimal price in each period will be a simple markup over that period's nominal marginal cost, subject to the limits of j's information set:

$$P_t(j) = \left(\frac{\varepsilon}{\varepsilon-1}\right) E_t(j)\left[P_t MC_t(j)\right] \tag{3.11}$$

3.3.3 Market clearing

All markets clear each period, so that:

$$Y_t(j) = C_t(j)\,\forall t, j \tag{3.12a}$$

$$H_t = \int H_t(j)\,dj\,\forall t \tag{3.12b}$$

which implies that aggregate output is given by:

$$Y_t = Z_t H_t^{1-\alpha} \tag{3.13}$$

where aggregate TFP, Z_t, combines individual firm productivities and distortions from relative prices and transitory shocks to relative demand:

$$Z_t \equiv \left(\int A_t(j)^{-(1+\eta)}\left(\frac{P_t(j)}{P_t}\right)^{-\varepsilon(1+\eta)} e^{-(1+\eta)v_{y,t}(j)}dj\right)^{-\frac{1}{1+\eta}} \tag{3.14}$$

[7] This obviously includes productivity shocks, but may be considered to also include shocks to the firm's marginal costs from other factors of production.

3.3.4 The central bank

To close the model, we assume that the central bank sets nominal interest rates according to the Taylor-like policy function:

$$i_t = \kappa_y E_t^{CB}[y_t] + \kappa_\pi E_t^{CB}[\pi_{t+1}] + \epsilon_{Mt} \tag{3.15}$$

where $E_t^{CB}[\cdot]$ is the mathematical expectation conditional on the central bank's information set in period t (defined below) and ϵ_{Mt} is a persistent, mean zero shock to monetary policy (specified below). Note that the component against inflation is against expected future inflation rather than current inflation, to provide a more accurate characterisation of modern central banking practice.

3.3.5 Stochastic processes

The underlying state of the economy, $\boldsymbol{x}_t$, therefore contains four aggregate shocks:

$$\boldsymbol{x}_t \equiv \begin{bmatrix} \epsilon_{At} & \epsilon_{Ct} & \epsilon_{Ht} & \epsilon_{Mt} \end{bmatrix}' \tag{3.16}$$

Of these, the shocks to productivity (ϵ_{At}) and the disutility of labour (ϵ_{Ht}) are pure supply shocks as they enter the model only through firms' marginal costs (although note that the latter acts via higher real wages); the shock to monetary policy (ϵ_{Mt}) is a pure demand shock as it only enters through the IS (Eular) relation; and the shock to the utility of consumption (ϵ_{Ct}) has both supply and demand aspects in that it affects both the spending/saving decision and the labour supply.

The underlying state is assumed to follow an AR(1) process:

$$\boldsymbol{x}_t = A\boldsymbol{x}_{t-1} + \boldsymbol{u}_t \tag{3.17}$$

where $\boldsymbol{u}_t$ is a vector of period-t innovations identically and independently distributed as $N(\mathbf{0}, I)$ and A is a matrix of fixed and commonly known parameters.

Idiosyncratic shocks to firms' productivity ($v_{a,t}(j)$), wages ($v_{w,t}(j)$) and demand ($v_{y,t}(j)$) are gathered together as

$$\boldsymbol{v}_t(j) \equiv \begin{bmatrix} v_{a,t}(j) \\ v_{w,t}(j) \\ v_{y,t}(j) \end{bmatrix} \tag{3.18}$$

The vector $\boldsymbol{v}_t(j)$ is entirely transitory, fully independent and jointly distributed as $N(\mathbf{0}, \sigma_v^2 I)$.

3.3.6 Firms' (linearised) marginal costs

We show in the appendix that firms' (linearised) real marginal costs are given by

$$mc_t(j) = \left(\eta + \frac{1}{\sigma} + \frac{1+\eta}{\psi}\right) y_t + \eta v_{y,t}(j) - \eta\varepsilon\left(p_t(j) - p_t\right) + \omega_t(j) \tag{3.19}$$

where $\omega_t(j)$ is the combined supply shock for firm j in period t:

$$\omega_t(j) \equiv B\boldsymbol{x}_t + Q\boldsymbol{v}_t(j) \tag{3.20a}$$

$$B = \begin{bmatrix} -(1+\eta)\left(1+\frac{1}{\psi}\right) & -1 & 1 & 0 \end{bmatrix} \tag{3.20b}$$

$$Q = \begin{bmatrix} -(1+\eta) & 1 & 0 \end{bmatrix} \tag{3.20c}$$

From this, we define $\dddot{mc}_t(j)$ to be a *partial average* of the firm's real marginal cost: the real marginal cost that firm j would incur without idiosyncratic demand shocks and if called upon to produce the average quantity (i.e. if $v_{y,t}(j) = 0$ and $y_t(j) = y_t$), but still with idiosyncratic supply shocks:

$$\dddot{mc}_t(j) \equiv \left(\eta + \frac{1}{\sigma} + \frac{1+\eta}{\psi}\right) y_t + \omega_t(j) \tag{3.21}$$

Finally, we define $\overline{mc}_t$ as the (true) average real marginal cost. That is, the real marginal cost a firm would incur if facing the average demand and experiencing the average supply shock (i.e. if producing the average quantity of output *and* experiencing no idiosyncratic shocks):

$$\overline{mc}_t \equiv \left(\eta + \frac{1}{\sigma} + \frac{1+\eta}{\psi}\right) y_t + B\boldsymbol{x}_t \tag{3.22}$$

3.3.7 Information and the network structure

Households and the central bank are assumed to possess complete information,[8] so that:

$$E_t^{CB}\left[X_{t-s}\right] = E_t^{HH}\left[X_{t-s}\right] = X_{t-s}\,\forall s \geq 0$$

[8] An arguably more plausible (and certainly more interesting) scenario would be to restrict the household and central bank to less than complete information so that dynamics might arise from their higher-order expectations of each others' and firms' beliefs. Exploration of such a setting is held for future research.

for any variable X_t, and to update their beliefs rationally, so that:

$$E_t^{CB}\left[X_{t+s}\right] = E_t^{HH}\left[X_{t+s}\right] = E\left[X_{t+s}|\Omega_t\right] \; \forall s \geq 1$$

where Ω_t is the set of all possible information available in period t.

Firms possess only incomplete information. We make an assumption of joint rationality, in the sense of Nimark (2008), so that all firms know the structure and the coefficients of the solution presented in section 3.3.9 below and firm j's information set evolves as:

$$\mathcal{I}_0\left(j\right) = \left\{F, G_1, G_2, G_3, \boldsymbol{\gamma}_p, \boldsymbol{\gamma}_y, \boldsymbol{\delta}_y, \sigma_v^2/\sigma_u^2, \Phi\right\} \tag{3.23a}$$

$$\mathcal{I}_t\left(j\right) = \left\{\mathcal{I}_{t-1}\left(j\right), \boldsymbol{s}_t\left(j\right)\right\} \tag{3.23b}$$

with $\boldsymbol{s}_t\left(j\right)$ being their set of public, private and social signals in period t.

For private signals, we will always maintain that firms observe their current-period supply shock $\left(\omega_t\left(j\right)\right)$ and the quantity of goods they produced and sold in the previous period $\left(y_{t-1}\left(j\right)\right)$. For public signals, we will variously suppose that all firms receive common, but imperfect, signals of the previous period's price level and real GDP.

In addition to these, we suppose that firms receive *social* signals by observing the previous-period prices set by individual competitors across an observation network characterised by link distribution Φ. That is, $\boldsymbol{s}_t\left(j\right)$ includes the set $\boldsymbol{g}_{t-1}\left(\boldsymbol{\delta}_{t-1}\left(j\right)\right)$, where $\boldsymbol{\delta}_{t-1}\left(j\right)$ is firm j's period-t draw from Φ, mapping them onto the index of a subset of firms that reset their price in period $t-1$.

For simplicity, the network is assumed to be effectively destroyed and redrawn each period. It therefore satisfies assumption 1 of chapter 2: that the network is opaque. The network is further assumed to be asymptotically irregular, in that the distribution Φ satisfies assumption 2 of chapter 2 so that, by proposition 2, we can define a vector of *network shocks* as:

$$\boldsymbol{z}_t \equiv \begin{bmatrix} {}^{\{1\}}\widetilde{\boldsymbol{v}}_t \\ {}^{\{2\}}\widetilde{\boldsymbol{v}}_t \\ {}^{\{3\}}\widetilde{\boldsymbol{v}}_t \\ \vdots \end{bmatrix} \sim N\left(\mathbf{0}, \Sigma_{zz}\right) \qquad \Sigma_{zz} = \begin{bmatrix} \Sigma_{\widetilde{v}\widetilde{v}}^{\{1\}} & \Sigma_{\widetilde{v}\widetilde{v}}^{\{1\}} & \Sigma_{\widetilde{v}\widetilde{v}}^{\{1\}} & \cdots \\ \Sigma_{\widetilde{v}\widetilde{v}}^{\{1\}} & \Sigma_{\widetilde{v}\widetilde{v}}^{\{2\}} & \Sigma_{\widetilde{v}\widetilde{v}}^{\{2\}} & \cdots \\ \Sigma_{\widetilde{v}\widetilde{v}}^{\{1\}} & \Sigma_{\widetilde{v}\widetilde{v}}^{\{2\}} & \Sigma_{\widetilde{v}\widetilde{v}}^{\{3\}} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{3.24a}$$

where

$$ {}^{\{q\}}\widetilde{\boldsymbol{v}}_t = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{v}_t \left(\underbrace{\delta_t(\cdots(\delta_t}_{q}(i))) \right) \tag{3.24b} $$

$$ \Sigma_{\widetilde{v}\widetilde{v}}^{\{q\}} = \left(1 - (1-\zeta^*)^q\right) \Sigma_{vv} \tag{3.24c} $$

and where $\zeta^* \equiv \lim_{N\to\infty} \sum_{i=1}^{N} \phi_N(i)^2$ is the degree of irregularity in the network, with $\zeta^* = 0$ indicating an asymptotically uniform link distribution and $\zeta^* = 1$ indicating a degenerate distribution.

The precise origin of the network (in this case, the origin of Φ) is largely left for future research and is here assumed to be exogenous. However, we do make two points of note in this regard.

First, we observe that social optimality dictates that there be at least some degree of irregularity in the network. When firms have incomplete information and face strategic complementarities, there is an incentive for them to overcome the coordination problem by all observing the same signal (i.e. have the distribution be degenerate, with everybody observing the price of one particular firm with probability 1). However, doing so would maximise the variance of the network shocks (which, recall, are aggregated idiosyncratic shocks). A social planner constrained to only choosing the distribution Φ would therefore face a trade-off.[9] A uniform distribution would minimise aggregate volatility and so raise welfare, but a degenerate distribution would solve the price-coordination problem. Exactly how non-uniform a distribution (i.e. how irregular a network) would be optimal would therefore be determined by the strength of the strategic complementarity. A higher coefficient against the average price in the individual firm's optimal decision rule would make the coordination problem more important and so result in a more asymmetric optimal distribution.

Second, we conjecture that the constrained social planner's optimal distribution will be a stable equilibrium. That is, conditional on all other firms drawing the identity of their observees from the distribution Φ, it will be optimal for any individual

[9] Of course, an unconstrained social planner would not face the informational or coordination problems in the first place.

firm to do the same. In appendix 3.B we provide a proof that this is true for a simple setting where a continuum of firms set static (i.e. one-period) prices and decide on the probabilities to assign to each of two potential observees.

3.3.8 Timing

Each period is divided into two phases, with the full gamut of innovations for the period occurring as the period begins.

1. In phase one, firms observe their public and private signals and the previous-period prices of a number of their competitors. Using this information, they set their prices for the current period.
2. In phase two, the central bank and the representative household, both of whom have full information, set the market-clearing interest rate and average nominal wage. The household reveals the quantity demanded from each firm at the given prices, firms discover their current-period marginal cost and produce the goods. The household consumes the goods entirely.

The key point of this timing arrangement is to ensure that while firms receive signals and know in advance that markets will clear, they cannot know exactly what their marginal cost will be, even in the current period (because they do not know what demand will be), when setting their prices.

3.3.9 Characterising the model solution

As per theorem 1 from chapter 2, the opaque nature of firms' observation network implies that their network learning problem will involve the simple-average expectation (because of firms' concern with the simple-average price) and a sequence of weighted average expectations because of the irregularity and opacity of the network. The full hierarchy of firms' expectations regarding $\boldsymbol{x}_t$ will therefore be defined recursively as:

$$X_t \equiv \begin{bmatrix} \boldsymbol{x}_t \\ \overline{E}_t\left[X_t\right] \\ {}^{\{1\}}\widetilde{E}_t\left[X_t\right] \\ {}^{\{2\}}\widetilde{E}_t\left[X_t\right] \\ \vdots \end{bmatrix} \tag{3.25}$$

and will follow an ARMA(1,1) process given by:

$$X_t = FX_{t-1} + G_1\boldsymbol{u}_t + G_2\boldsymbol{z}_t + G_3\boldsymbol{e}_t + G_4\boldsymbol{z}_{t-1} \tag{3.26}$$

On the demand side of the economy, we have the familiar linearised household Eular equation (3.4) and the central bank's policy function (3.15):

$$\begin{aligned} y_t &= E_t\left[y_{t+1}\right] - \sigma E_t\left[i_t - \pi_{t+1}\right] + \sigma\left(\epsilon_{Ct} - E_t\left[\epsilon_{Ct+1}\right]\right) \\ i_t &= \kappa_y y_t + \kappa_\pi E_t\left[\pi_{t+1}\right] + \epsilon_{Mt} \end{aligned}$$

where $E_t\left[\cdot\right]$ is the full-information expectation formed by the representative household and the central bank. We combine these two to write simply

$$y_t = \frac{1}{1+\sigma\kappa_y}E_t\left[y_{t+1} - \sigma\left(\kappa_\pi - 1\right)\left(p_{t+1} - p_t\right)\right] + \boldsymbol{\mu}_y'\boldsymbol{x}_t \tag{3.27a}$$

where $\boldsymbol{\mu}_y$ is given by

$$\boldsymbol{\mu}_y' = \frac{\sigma}{1+\sigma\kappa_y}\left(\begin{bmatrix}0 & 1 & 0 & -1\end{bmatrix} - \begin{bmatrix}0 & 1 & 0 & 0\end{bmatrix}A\right) \tag{3.27b}$$

On the supply side of the economy, with firms free to adjust their prices in every period, the linearised expression for their decision rule will be

$$p_t\left(j\right) = E_t\left(j\right)\left[p_t + \frac{1}{1+\varepsilon\eta}\dddot{mc}_t\left(j\right)\right] \tag{3.28}$$

where $\dddot{mc}_t\left(j\right)$ is the real marginal cost faced by a firm with an idiosyncratic demand shock of zero that is called upon to produce the average quantity of goods.

With firms observing $\omega_t\left(j\right)$ directly and defining $\chi \equiv \left(\frac{1}{1+\varepsilon\eta}\right)\left(\eta + \frac{1}{\sigma} + \frac{1+\eta}{\psi}\right)$, we show in the appendix that the aggregate price level may then be written as

$$p_t = \left(\frac{1}{1+\varepsilon\eta}\right)B\boldsymbol{x}_t + \sum_{k=0}^{\infty}\xi^k\overline{E}_t^{(k+1)}\begin{bmatrix}\left(\chi\boldsymbol{\mu}_y' + \xi\left(\frac{1}{1+\varepsilon\eta}\right)B\right)\boldsymbol{x}_t \\ +\chi\left(\frac{1}{1+\sigma\kappa_y}\right)\left(y_{t+1} + \sigma\left(1-\kappa_\pi\right)p_{t+1}\right)\end{bmatrix} \tag{3.29a}$$

where

$$\xi \equiv 1 - \left(\frac{1}{1+\varepsilon\eta}\right)\left(\eta + \frac{1}{\sigma} + \frac{1+\eta}{\psi}\right)\left(\frac{\sigma}{1+\sigma\kappa_y}\right)(1-\kappa_\pi) \tag{3.29b}$$

which is to say that the aggregate price level is a function of the current period average supply shock and the full hierarchy of average expectations regarding (a) the current period average supply shock; (b) the next period's aggregate demand; and (c) the next period's price level. Note that in order for the aggregate price to be well defined, we require $\xi \in (0,1)$ and that this requires $\kappa_\pi < 1$.[10]

The reduced-form solution

The solution may then be found by the method of undetermined coefficients. That is, in addition to the law of motion for firms' beliefs (3.26), we next posit (and verify) reduced-form solutions of the following form:

$$p_t = \boldsymbol{\gamma}_p' X_t \tag{3.30a}$$
$$y_t = \boldsymbol{\gamma}_y' X_t + \boldsymbol{\delta}_y' \boldsymbol{z}_t \tag{3.30b}$$

and similarly suppose expressions for the real wage and hours worked as:

$$w_t - p_t = \boldsymbol{\gamma}_\varpi' X_t + \boldsymbol{\delta}_\varpi' \boldsymbol{z}_t \tag{3.30c}$$
$$h_t = \boldsymbol{\gamma}_h' X_t + \boldsymbol{\delta}_h' \boldsymbol{z}_t \tag{3.30d}$$

Note that current-period network shocks do not appear in the reduced-form expression for the price level because firms do not contemporaneously observe any signals that rely upon them. Network shocks *do* appear in the expressions for all real variables, however, because of the assumptions that (a) the representative household and central bank have full information; and (b) markets clear.

[10] If the central bank's reaction function (3.15) were to also include a term in *current* inflation as well (for example, $i_t = \kappa_y y_t + \kappa_{\pi_0}\pi_t + \kappa_{\pi_1} E_t[\pi_{t+1}] + \epsilon_{Mt}$) then the final term in the expression for ξ would be $(1 + \kappa_{\pi_0} - \kappa_{\pi_1})$, therefore allowing a stronger response to expected next-period inflation. We have not done so here in order to ensure that firms' prices are not backward looking.

In appendix 3.A.5, we confirm the reduced-form solution and derive the following conditions under which it holds:

$$\begin{aligned}\boldsymbol{\gamma}_p' = &\left(BS_x + \begin{pmatrix} \left(\chi \boldsymbol{\mu}_y' + \xi B\right) S_x \\ +\chi \left(\frac{1}{1+\sigma\kappa_y}\right) \boldsymbol{\gamma}_y' F \end{pmatrix} \left(I - \xi T_s\right)^{-1} T_s \right) \\ &\times \left(I - (1-\xi) F \left(I - \xi T_s\right)^{-1} T_s\right)^{-1} \end{aligned} \tag{3.31a}$$

$$\boldsymbol{\gamma}_y' = \left(\boldsymbol{\mu}_y' S_x + \frac{\sigma}{1+\sigma\kappa_y} \left(1-\kappa_\pi\right) \boldsymbol{\gamma}_p' \left(F - I\right) \right) \left(I - \frac{1}{1+\sigma\kappa_y} F \right)^{-1} \tag{3.31b}$$

$$\boldsymbol{\delta}_y' = \frac{1}{1+\sigma\kappa_y} \left(\boldsymbol{\gamma}_y' + \sigma \left(1-\kappa_\pi\right) \boldsymbol{\gamma}_p' \right) G_3 \tag{3.31c}$$

$$\boldsymbol{\gamma}_\varpi' = \left(\frac{1}{\sigma} + \frac{1+\eta}{\psi} \right) \boldsymbol{\gamma}_y' + \begin{bmatrix} -\frac{1+\eta}{\psi} & -1 & 1 & 0 \end{bmatrix} S_x \tag{3.31d}$$

$$\boldsymbol{\delta}_\varpi' = \left(\frac{1}{\sigma} + \frac{1+\eta}{\psi} \right) \boldsymbol{\delta}_y' \tag{3.31e}$$

$$\boldsymbol{\gamma}_h' = (1+\eta) \boldsymbol{\gamma}_y' - (1+\eta) \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} S_x \tag{3.31f}$$

$$\boldsymbol{\delta}_h' = (1+\eta) \boldsymbol{\delta}_y' \tag{3.31g}$$

In appendix 3.A.6, we characterise firms' signal vectors in terms of the state vector of interest and show that theorem 1 from the previous chapter applies, so that the posited state law of motion (3.26) is indeed correct.

Determinacy of the solution

Given the assumptions that the central bank (a) does not respond to current inflation and (b) responds to next-period inflations with a coefficient less than one, the reader may be concerned about the determinacy of the model as, on the face of it, this violates the Taylor Principle that monetary authorities respond by more than one-for-one to inflation. However, it should be noted that the Taylor principle relates to settings where firms have access to full information, which does not apply here.

Instead, uniqueness of the solution is established by, first, the method of undetermined coefficients pinning down expressions for $\boldsymbol{\gamma}_p$, $\boldsymbol{\gamma}_y$ and $\boldsymbol{\delta}_y$ (so that p_t and y_t are linear functions of firms' hierarchy of expectations and network shocks only) and, second, the recursive projection of firms' expectations onto the complete history of their observables via a Kalman filter to pin down the law of motion for X_t under incomplete information (i.e. to demonstrate that X_t is a function of fundamental shocks only) in an extension of the methodology of Nimark (2008, 2011a).

Unlike the matrix decomposition methods of Blanchard and Kahn (1980) or Klein (2000), the solution methodology here does not identify the range of parameters for which a unique solution exists. In particular, the assumption that $\kappa_\pi \in (0, 1)$ is *not* a requirement for determinacy. It is instead a necessary condition for firms to place decreasing weight on higher-order expectations so that the full solution may be well approximated by simulating only a finite subset of X_t. Indeed, for the baseline parameterisation outlined below, convergence to a stable solution appears to require values of $\kappa_\pi \geq 0.5$.

3.3.10 Finding the solution

Finding the true solution to the model requires working with expectations of infinite order, which cannot be handled in practice. However, so long as ξ (3.29b) lies between zero and unity, the model places decreasing weight on higher order expectations (a weight of ξ^k is applied to the average k-th order expectation), an arbitrarily accurate approximation of the solution may be found by truncating firms' expectation hierarchy at an upper limit, k^*, of the number of orders to include. Recall, from the previous chapter, that the recursive nature of agents' (here firms') learning and the AR process for the underlying state ensures that decreasing weight is also applied to higher weighted average expectations, thus allowing us to also impose an upper limit, p^*, on the number of compound expectations to include.

For a given set of parameters and chosen values for k^* and p^*, the solution is obtained by finding the fixed point of the system (3.26), (3.30) and (3.31).

In practice, given the size of state vector, X_t, care must be taken to avoid the numerical instability issues described in the previous chapter. Since the solution here involves finding the simultaneous fixed points of *three* systems of equations – the Kalman filter and the state law of motion from the previous chapter; and the macroeconomic coefficients detailed above – the root loop of the solution algorithm must be expanded to:

repeat
  Update the filter by one step, using equation (2.13)
  Update the law of motion by one step, using equation (2.14)
  Update the macroeconomic coefficients by one step, using equation (3.31)
until all three converge

3.4 Simulation

Table 3.1 lists baseline parameters for the simulation presented below. Most values should be uncontroversial, but a number of points bear highlighting. First, note that each aggregate state variable (i.e. each element of $\boldsymbol{x}_t$) is assumed to follow an independent AR(1) process, with interaction between them occurring only within the broader model. Given the simplicity of the model, this is arguably too restrictive an assumption – for example, a monetary policy shock (ϵ_{Mt}) could plausibly affect commodity prices which in the current model would appear as an aggregate shock to marginal costs (ϵ_{At}), implying that a realistic simulation should permit the two to covary – but is nevertheless made in order to ensure that the firms' learning problem is as easy as possible and thus to avoid any bias to the strength of our results.

Parameter	Value	Description
A	$0.6I_4$	The AR(1) transition matrix for the underlying state
P	I_4	The map from true aggregate shocks to the underlying state
Σ_{uu}	I_4	The variance-covariance matrix for true aggregate shocks
Σ_{vv}	$5I_3$	The variance of idiosyncratic shocks
κ_y	0.5	The CB's coefficient against current real GDP
κ_π	0.5	The CB's coefficient against expected next-period inflation
ε	3	The own-price elasticity of demand
σ	1/3	The elasticity of intertemporal substitution
ψ	1.5	The Frisch elasticity of labour supply
η	0.5	The elasticity of marginal cost ($\equiv \frac{\alpha}{1-\alpha}$, $\alpha = 0.333$)
q	2	The number of competitors observed by each firm
ζ^*	0.2	The degree of irregularity in firms' observation network (0 = uniform, 1 = degenerate)
ξ	0.743	The relative importance of higher-order expectations (implied by parameters above)

Table 3.1: Baseline parameterisation

Next, because we wish to focus on the relative effect of idiosyncratic shocks, we normalise the variance of all true aggregate shocks to 1 and vary only the idiosyncratic shock variance (with a baseline value of 5). The degree of irregularity in the firms' observation network, ζ^*, represents a new concept in this paper and cannot be known with any certainty. For our baseline, we set it to 0.2, which is the value that would emerge from a power-law distribution with a shape parameter of $\gamma = 1.5$. Note that with these values, the implied variance of the aggregated idiosyncratic shocks, ${}^{\{1\}}\widetilde{\boldsymbol{v}}_t$ and ${}^{\{2\}}\widetilde{\boldsymbol{v}}_t$, are 1.0 and 1.8 respectively.

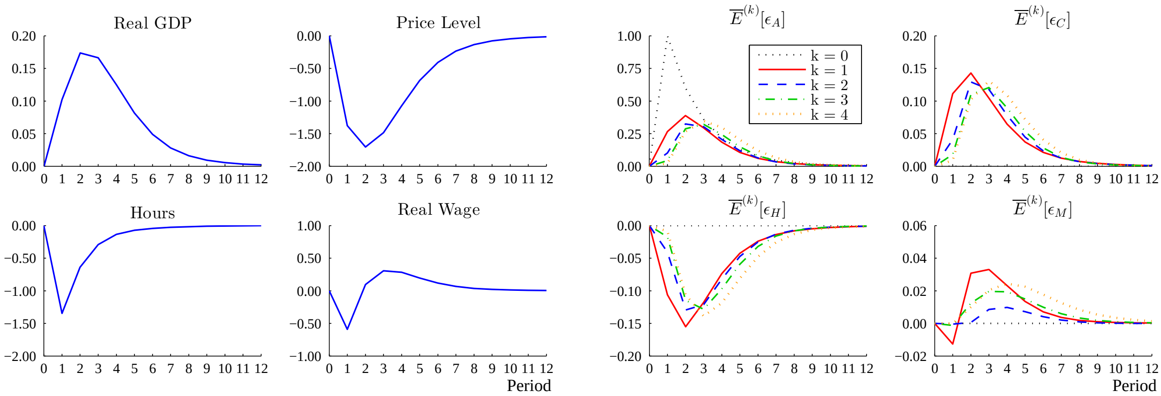
Finally, as a baseline, we suppose that firms observe their private signals (i.e. their current-period supply shock and their previous-period quantity demanded) and the previous-period prices of two competitors, but no public signals.

In everything that follows, period 0 denotes the period immediately prior to any shock occurring (the economy is invariably assumed to be in steady-state in period 0) and period 1 denotes the "on impact" period.

3.4.1 Responses to aggregate shocks

Baseline responses

Figures 3.1 to 3.4 plot impulse responses following shocks of one standard deviation to each of the four underlying state variables under the baseline parameterisation.

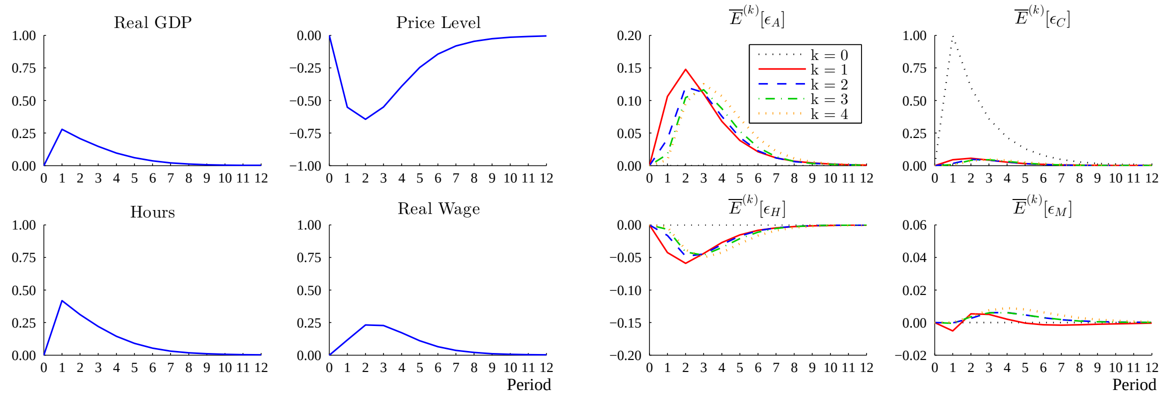


(a) Aggregate variables

(b) Hierarchies of simple-average expectations

Figure 3.1: IRFs following a one s.d. shock to firms' aggregate productivity

In each case, firms cannot be certain about the origin of the shock and subsequently believe that a combination of all four aggregate shocks (and idiosyncratic shocks) have occurred. The impulse responses of macroeconomic variables are generally of the expected sign, although for the price level following a shock to the utility of consumption, the supply-side effect (through the labour supply) dominates the demand side effect (through the Eular equation), leading to a price decline.



(a) Aggregate variables

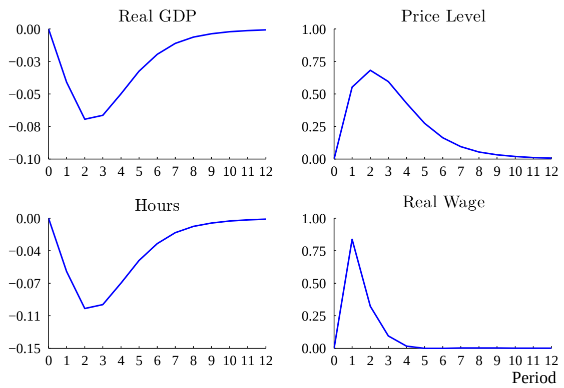
(b) Hierarchies of simple-average expectations

Figure 3.2: IRFs following a one s.d. shock to the utility of consumption

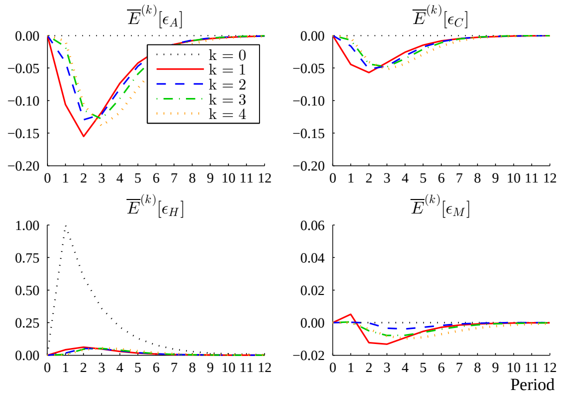
Note, too that firms' beliefs respond on impact for the first three shocks, but with a one period lag for the monetary policy shock. This is because the three shocks with supply-side effects are observable immediately through firms' marginal costs, but the latter affects only demand, which firms do not observe at the time of setting their prices.

The lag in observability of demand also explains why the magnitude of the response of real GDP is so much larger than that for the price level following a monetary policy shock. In contrast, for shocks to the three variables with supply-side effects, the magnitude of price changes is greater than that for real GDP, reflecting the immediacy of firms' signals and their freedom to adjust their prices every period.

Varying the standard parameters of the system induces expected changes in the impulse responses. Instead, we next focus on varying the three parameters that affect firms' network learning.

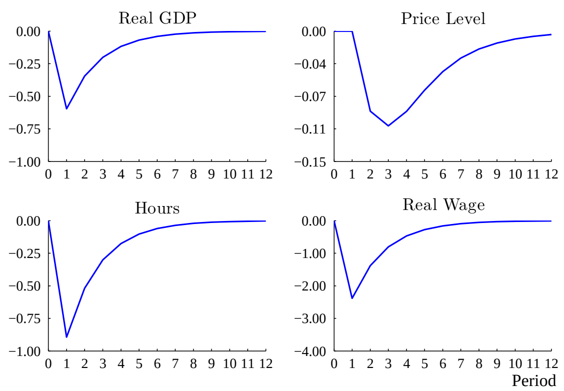


(a) Aggregate variables

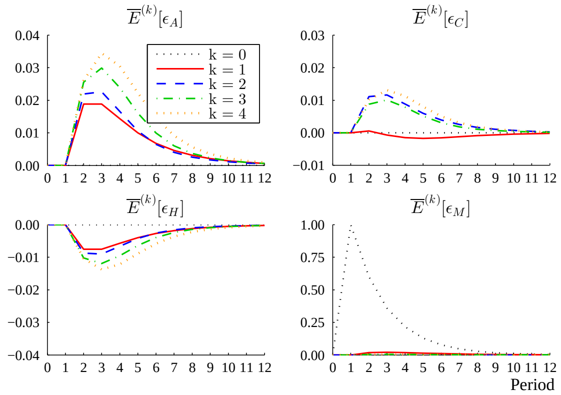


(b) Hierarchies of simple-average expectations

Figure 3.3: IRFs following a one s.d. shock to the disutility of labour

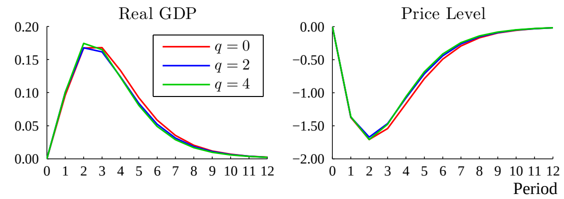


(a) Aggregate variables

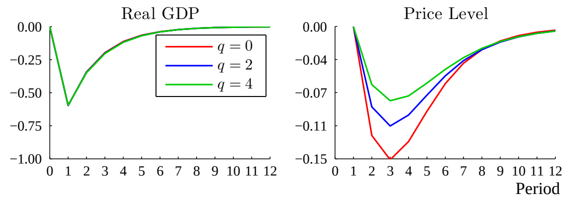


(b) Hierarchies of simple-average expectations

Figure 3.4: IRFs following a one s.d. monetary policy shock



(a) Following a one s.d. shock to aggregate productivity



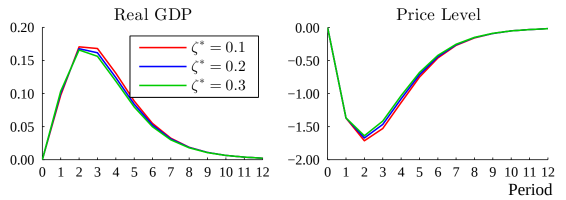
(b) Following a one s.d. monetary policy shock

Figure 3.5: IRFs for various numbers of competitors observed

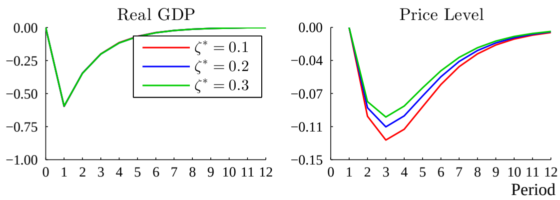
Varying the number of competitors observed

Figure 3.5 shows impulse responses for real GDP and the price level following shocks to aggregate productivity or monetary policy for a variety of numbers of competitors observed. The addition of extra observees tightens all impulse responses for all shocks, but has a marked effect in reducing the magnitude (but not the persistence) of deviations of the price level from trend as individual firms are better able to form estimates of the average price.

Varying the network irregularity



(a) Following a one s.d. shock to aggregate productivity



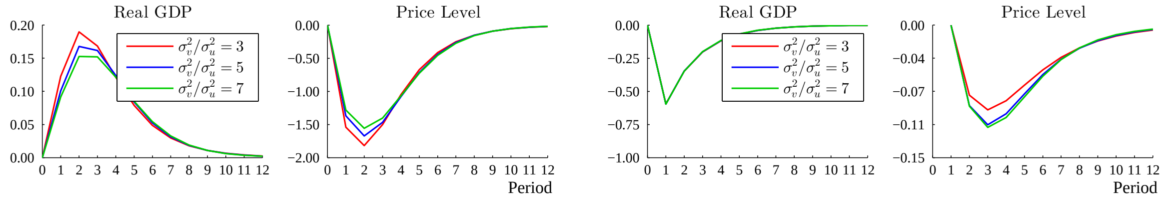
(b) Following a one s.d. monetary policy shock

Figure 3.6: IRFs for various degrees of network irregularity

Figure 3.6 plots the same impulse responses for a variety of values of ζ^*. As with increasing the number of competitors observed, making the observation network more irregular tighten the IRFs slightly for the supply-side shock, but has its largest effect in the IRF for the price level following a monetary policy shock. Differently to the previous case, however, increasing ζ^* not only lowers the magnitude of the deviation, but lowers its persistence as well. This is because of the subsequent increased ability of observations of competitors' prices to act as a common signal, thereby representing a herding device.

Varying the relative signal variance

Figure 3.7 plots the same impulse responses for different levels of variance in firms' idiosyncratic shocks.



(a) Following a one s.d. shock to aggregate productivity

(b) Following a one s.d. monetary policy shock

Figure 3.7: IRFs for different levels of relative signal variance

3.4.2 Responses to network shocks

To say that the economy experiences a network shock, we mean that a full suite of idiosyncratic shocks occur whose combined *effect* does not "wash out" in aggregate. In the context of the current model, this means that the more prominent firms happen to experience shocks in one direction (say, for example, a cost increase causing their prices to go up) while more obscure firms experience shocks in the opposite direction. On average across all firms, these shocks cancel out almost surely, but because the firms note each others' prices in an observation network that is irregular, a typical firm is more likely to observe a competitor's price rise than fall. From this, they may conclude (a) that average costs have indeed gone up; or (b) that, at the least, other businesses will *believe* that they have gone up. In either event, it becomes rational for the typical firm to increase their own price too, even in the absence of other signals suggesting such an action.

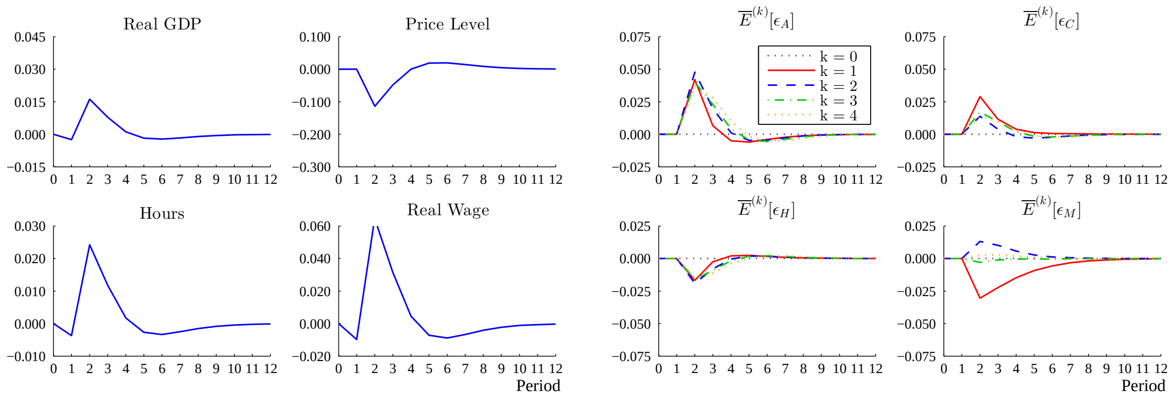
Baseline responses

Figures 3.8 to 3.10 plot impulse responses following shocks of one standard deviation to each of the three innovations in ${}^{\{1\}}\widetilde{\boldsymbol{v}}_t$ and the corresponding conditionally expected value in the other network shocks under the baseline parameterisation.[11]

In general, despite network shocks having the same variance as underlying state shocks under the baseline parameterisation, their effects are roughly one order of magnitude smaller and, while persistent, less so than for aggregate shocks. This

[11] Recall that corollary 2 following proposition 2 in the previous chapter gives us that $E\left[{}^{\{q\}}\widetilde{v}_t \,|\, {}^{\{1\}}\widetilde{v}_t = \boldsymbol{a}\right] = \boldsymbol{a}\;\forall q \geq 2$.

latter point is not surprising given that idiosyncratic shocks are purely transitory, so that all persistence demonstrated here derives from real rigidities evoked from firms' learning and herding behaviour.



(a) Aggregate variables

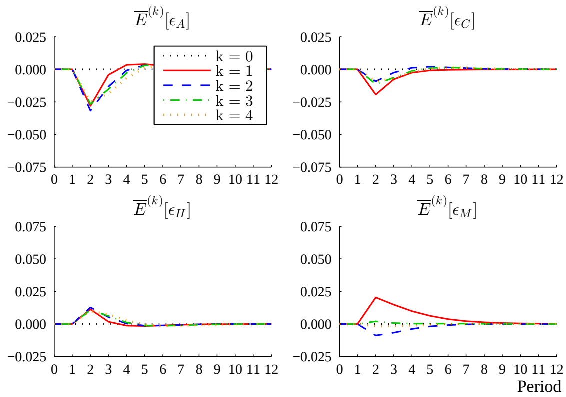
(b) Hierarchies of simple-average expectations

Figure 3.8: IRFs following a one s.d. shock to $\widetilde{v}_{A,t}$

For idiosyncratic shocks to prominent firms' productivity, the direction of deviations for real GDP and the price level are the same as for an aggregate productivity shock, but since aggregate productivity does not actually increase, this is only achieved through a temporary boost in the demand for labour by firms. This is because the average firm does not experience any change in their productivity, but having seen price falls at their competitors, believes that average productivity has increased. Concluding that *average* prices will fall with the productivity rise, firms then lower their individual prices (making their beliefs self-fulfilling). The lower price level prompts greater demand and, without any actual increase in average productivity, this is met through an increased demand for labour.

Idiosyncratic shocks to prominent firms' wage bargaining have almost identical, but inverse, effects as for idiosyncratic productivity shocks, as because both are visible to firms only through their observation of the combined supply-side shock. Firms' inability to fully differentiate between the two explains why expectations of both aggregate productivity and aggregate labour-supply shocks move in response.

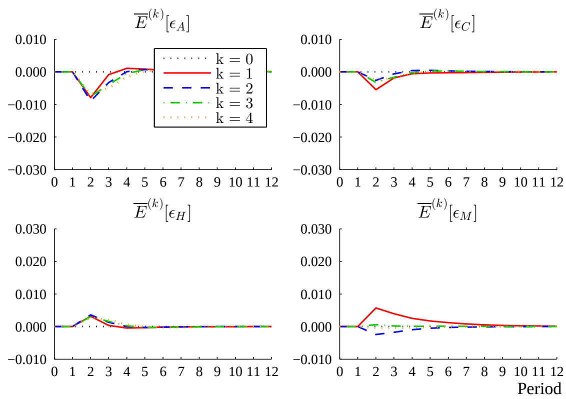
(a) Aggregate variables



(b) Hierarchies of simple-average expectations

Figure 3.9: IRFs following a one s.d. shock to $\widetilde{v}_{W,t}$

(a) Aggregate variables

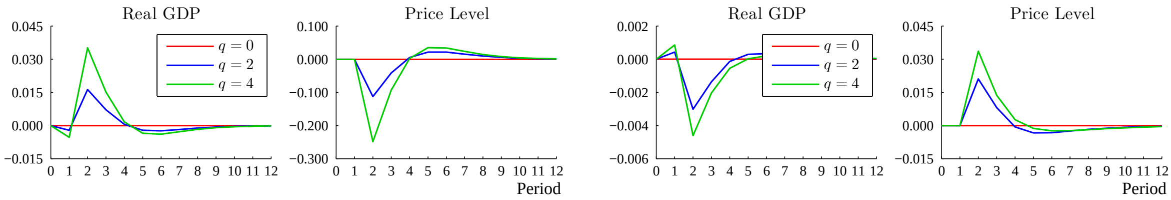


(b) Hierarchies of simple-average expectations

Figure 3.10: IRFs following a one s.d. shock to $\widetilde{v}_{Y,t}$

Interestingly, real GDP falls following a positive demand shock among more visible firms. The mistaken perception of extra aggregate demand causes firms to raise their prices in anticipation and this actually causes demand to fall.

Varying the number of competitors observed

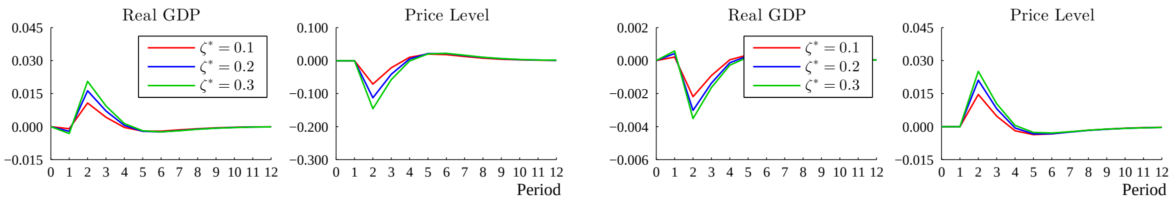


(a) Following a one s.d. shock to $\widetilde{v}_{A,t}$ (b) Following a one s.d. shock to $\widetilde{v}_{Y,t}$

Figure 3.11: IRFs for various numbers of competitors observed

Figure 3.11 shows impulse responses for real GDP and the price level following shocks to prominent firms' productivity and demand for a variety of numbers of competitors observed. When firms observe no competitors' prices, there is no effect on aggregate variables, but as the number of observed competitors increases, the magnitude of the aggregate response correspondingly rises.

Varying the network irregularity



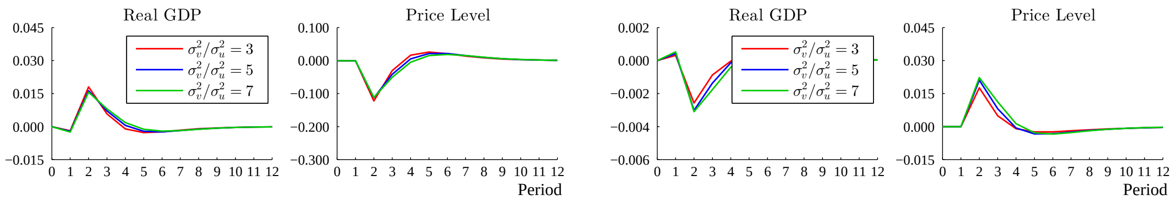
(a) Following a one s.d. shock to $\widetilde{v}_{A,t}$ (b) Following a one s.d. shock to $\widetilde{v}_{Y,t}$

Figure 3.12: IRFs for various degrees of network irregularity

Figure 3.12 shows impulse responses for real GDP and the price level following shocks to prominent firms' productivity and demand for a variety of values of ζ^*. For more irregular networks, observed prices are more concentrated among the more prominent firms, meaning that they serve as a better coordination device for herding.

This leads to a subsequent increase in the magnitude of aggregate deviations from trend following a network shock.

Varying the relative signal variance



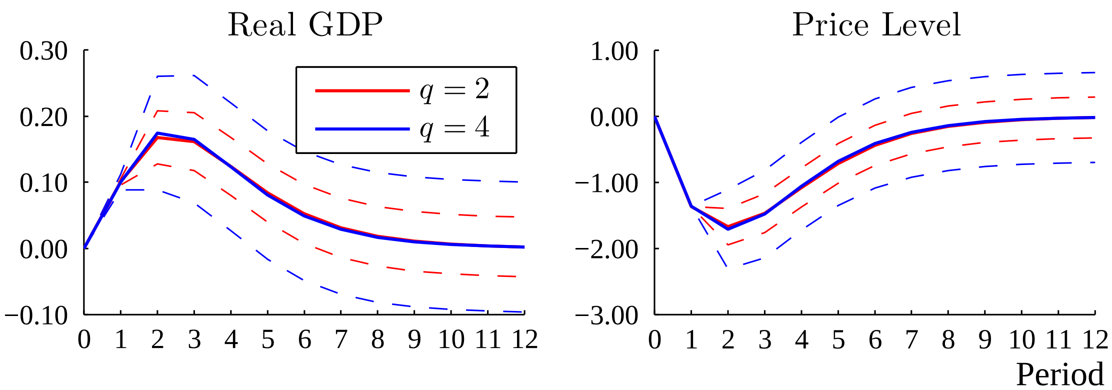
(a) Following a one s.d. shock to $\widetilde{v}_{A,t}$ (b) Following a one s.d. shock to $\widetilde{v}_{Y,t}$

Figure 3.13: IRFs for different levels of relative signal variance

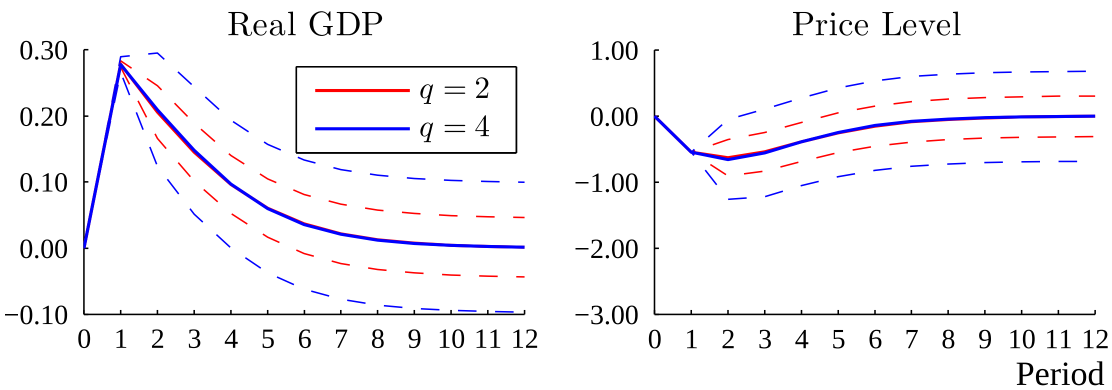
Figure 3.13 shows impulse responses for real GDP and the price level following shocks to prominent firms' productivity and demand for different levels of variance in firms' idiosyncratic shocks. Although the variance of network shocks is increasing in the variance of firms' idiosyncratic shocks, the increase in aggregate response is much more muted than for increases in network irregularity. This is because as firms' idiosyncratic variance increases, the *informational* value of individual firms' prices decreases, creating an offsetting effect.

3.4.3 Trade-offs in volatility

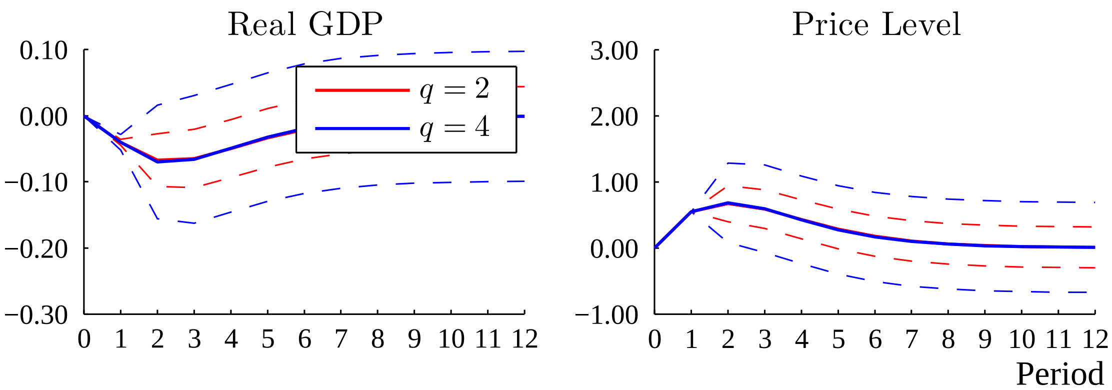
Figures 3.5 and 3.11 make clear that under the baseline parameterisation, there is a trade-off in aggregate volatility involved in firms increasing the number of competitors they observe. With no competitors observed, there are no network shocks and so no volatility from this source, but the magnitude of deviations following aggregate shocks – particularly monetary shocks – is correspondingly higher. To illustrate this trade-off, figure 3.14 plots the *distribution* of impulse responses that would occur following each of the four aggregate shocks for various numbers of competitors observed if network shocks are free to occur while aggregate shocks are held to their expected paths. Dotted lines represent 2 s.d. bands for the distribution of impulse responses that would occur, conditional on the given path for the four aggregate



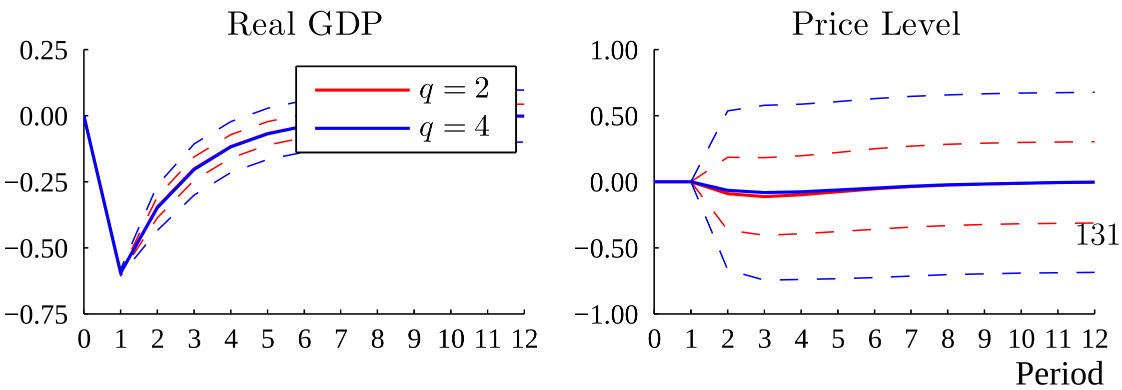
(a) Following a one s.d. shock to aggregate productivity



(b) Following a one s.d. shock to the utility of consumption



(c) Following a one s.d. shock to the disutility of labour



(d) Following a one s.d. monetary policy shock

Figure 3.14: IRFs for aggregate shocks with indicative bands for network shocks

shocks.[12] Increasing the number of competitors observed leads to a clear increase in volatility attributable to network shocks. In some cases, this may be enough to swamp the aggregate effects of the shock to the underlying state (although note that the IRFs shown here are for 1 s.d. shocks to underlying state variables, while the dashed lines represent 2 s.d. bands for the effects of network shocks).

Another way of exploring this is to perform a variance decomposition. Table 3.2 shows the share of *unconditional* variance in Real GDP and the Price Level that can be attributed to network shocks under the baseline parameterisation for different numbers of competitors observed and different degrees of asymmetry in the network. Although the share is quite low for real GDP, between 1% and 2% of unconditional

q \ ζ^*	0	0.1	0.2	0.3	0.4
0	0	0	0	0	0
1	0	0.01	0.01	0.02	0.03
2	0	0.03	0.06	0.10	0.14
3	0	0.06	0.15	0.24	0.32
4	0	0.12	0.30	0.46	0.61

(a) Real GDP

q \ ζ^*	0	0.1	0.2	0.3	0.4
0	0	0	0	0	0
1	0	0.02	0.05	0.08	0.11
2	0	0.08	0.21	0.35	0.51
3	0	0.19	0.50	0.85	1.22
4	0	0.37	0.99	1.67	2.36

(b) Price Level

Table 3.2: Share of unconditional variance attributable to network shocks (%)

volatility in the aggregate price level may be attributable to network shocks under the baseline parameterisation, even for quite low numbers of competitors observed. This is still relatively low, and so indicates that network shocks may be best described as explaining noise around deviations due to aggregate shocks. However, the share of volatility attributable to network shocks does increase notably as the number of observees or the degree of network irregularity increases. For $q = 4$ and $\zeta^* = 0.5$, network shocks contribute 3% of unconditional volatility in the price level.

[12] That is, were a researcher to simulate the economy described here by giving a persistent shock to one of the underlying aggregate state variables, holding all other aggregate state variables to zero and having a full gamut of idiosyncratic shocks occur in every period, the subsequent impulse responses would fall within the dashed lines 95% of the time.

3.5 Conclusion

This chapter has argued that firms set their prices while operating in an observation network, making use of competitors' prices to learn about the aggregate state of the economy. That firms operate in a network and that they do so in a model of imperfect common knowledge is motivated by the observation that when surveyed, a large fraction of firms across North America and Europe admit to looking to other firms in deciding both the timing and the magnitude of price changes and do so out of a desire to coordinate pricing changes with competitors.

When the observation network between firms is asymptotically irregular and the network is opaque, the results of chapter 2 apply, so that the effects of firms' mean zero idiosyncratic shocks do not "wash out" with aggregation. Instead, firms' hierarchy of average expectations will follow an ARMA(1,1) process with current and lagged network shocks (weighted sums of firms' idiosyncratic shocks) entering at an aggregate level. The recursive nature of agents' learning then implies that the aggregate effects of idiosyncratic shocks will be persistent, despite the individual agents' shocks being entirely transitory, with this persistence increasing in the degree of strategic complementarity, the asymmetry of the network and the persistence of any aggregate shocks.

These persistent aggregate effects therefore represent a network learning-based microfoundation for cost-push shocks, with the aggregate price level able to persistently deviate from it's long-run trend despite (a) the absence of any aggregate shocks to the economy; (b) firms being free to adjust their prices every period; and (c) network (i.e. idiosyncratic) shocks being purely transitory. Because firms may choose to observe the prices of other firms with whom they are are not direct competitors, this also represents a novel transmission mechanism for inflation across industries or geographies independent of it's path along production chains.

In contrast to the common assumption that idiosyncratic shocks cancel out in aggregation, the emergence of aggregate-level price changes based on short-lived idiosyncratic shocks is consistent with evidence garnered from a variety of observed panels of micro price changes. The level of aggregate volatility induced through network learning is increasing in the number of competitors observed, the asymmetry of the network and the relative variance of idiosyncratic shocks.

This model clearly calls for future work to estimate the parameters of the model – particularly q, ζ^* and σ_v^2/σ_u^2. While the obvious choice in this would be to pursue data on a panel of firms, the differential responses of aggregate variables predicted here following aggregate and idiosyncratic shocks may permit such an estimation even in the absence of individual firm data. The implications for optimal monetary policy are a second area of research that warrants further work. Just as previous work has suggested that monetary authorities focus their attention on the "stickiest" prices, it may also be necessary to focus on the most *visible* prices in the economy. Finally, further research into the origins of firms' observation networks would seem a fruitful area for exploration.

Appendix 3.A Derivation

This appendix provides a full derivation of the model of dynamic price setting with network learning presented in the text. We suppose that in steady state there are no shocks; technology and real output are constant; prices are constant; and all firms make the same decisions:

$$
\begin{aligned}
Y_t(j) &= Y^{ss} \\
A_t(j) &= A^{ss} \\
\Pi_t &= \Pi^{ss} = 1 \\
W_t(j) &= W^{ss}_{t+s} = W^{ss} \\
G_t(j) &= G^{ss} = P^{ss} \\
MC_t(j) &= MC^{ss} \\
Q_{t+s|t} &= Q^{ss} = 1
\end{aligned}
$$

We normalise $P^{ss} = 1$ and denote lower-case letters as log-deviations from the steady-state (e.g. $x_t \equiv \ln(X_t) - \ln(X^{ss})$).

3.A.1 The household and central bank

The derivation of the representative household's optimality conditions is entirely standard and therefore omitted.

Substituting the market-clearing requirements (3.12) into the household's Euler equation gives:

$$
e^{\epsilon_{Ct}} Y_t^{-\frac{1}{\sigma}} = \beta(1+i_t) E_t^{HH}\left[e^{\epsilon_{Ct+1}} Y_{t+1}^{-\frac{1}{\sigma}} \frac{1}{\Pi_{t+1}}\right]
$$

and linearising this gives:

$$
y_t = E_t^{HH}[y_{t+1}] - \sigma E_t^{HH}[i_t - \pi_{t+1}] + \sigma\left(\epsilon_{Ct} - E_t^{HH}[\epsilon_{Ct+1}]\right)
$$

Noting that the household and the central bank both have full information, we can therefore write:

$$
\begin{aligned}
y_t &= E_t[y_{t+1}] - \sigma(i_t - E_t[\pi_{t+1}]) + \sigma(\epsilon_{Ct} - E_t[\epsilon_{Ct+1}]) \\
i_t &= \kappa_y y_t + \kappa_\pi E_t[\pi_{t+1}] + \epsilon_{Mt}
\end{aligned}
$$

Combining these two then gives a linearised expression for aggregate demand in the economy:

$$y_t = \frac{1}{1+\sigma\kappa_y} E_t\left[y_{t+1} + \sigma\left(\kappa_\pi - 1\right)\left(p_{t+1} - p_t\right)\right] + \boldsymbol{\mu}_y' \boldsymbol{x}_t$$

with $\boldsymbol{\mu}_y$ given by

$$\boldsymbol{\mu}_y' = \frac{\sigma}{1+\sigma\kappa_y}\left(\begin{bmatrix}0 & 1 & 0 & -1\end{bmatrix} - \begin{bmatrix}0 & 1 & 0 & 0\end{bmatrix} A\right)$$

which is equation (3.27) in the main text.

3.A.2 The market-clearing (average) wage

The idiosyncratic nominal wage faced by firm j in period t is given by:

$$w_t\left(j\right) = w_t + v_{w,t}\left(j\right) \tag{3.32}$$

where $v_{w,t}\left(j\right)$ is a transitory, mean zero shock and w_t is the market-clearing nominal wage. To find this, we will substitute the aggregate demand for labour into the household's labour supply curve. We start by substituting the individual firm's production function (3.7) into the labour market clearing condition (3.12) to obtain:

$$H_t = \int \left(\frac{Y_t\left(j\right)}{A_t\left(j\right)}\right)^{1+\eta} dj$$

Further substituting in the firm's demand function (3.5) gives:

$$\begin{aligned} H_t &= \int \left(\frac{\left(\frac{P_t(j)}{P_t}\right)^{-\varepsilon} Y_t e^{v_{y,t}(j)}}{A_t\left(j\right)}\right)^{1+\eta} dj \\ &= Y_t^{1+\eta} \underbrace{\int A_t\left(j\right)^{-(1+\eta)} \left(\frac{P_t\left(j\right)}{P_t}\right)^{-\varepsilon(1+\eta)} e^{-(1+\eta)v_{y,t}(j)} dj}_{\equiv Z_t^{-(1+\eta)}} \end{aligned}$$

Rearranging (recall that $1+\eta = \frac{1}{1-\alpha}$), we arrive at:

$$Y_t = Z_t H_t^{1-\alpha}$$

which is equation (3.13) in the text. Substituting this into the household's labour supply FOC gives:

$$\frac{W_t}{P_t} = e^{\epsilon_{Ht}-\epsilon_{Ct}} Y_t^{\frac{1}{\sigma}+\frac{1+\eta}{\psi}} Z_t^{-\frac{1+\eta}{\psi}}$$

Linearising this gives:

$$w_t - p_t = \left(\frac{1}{\sigma} + \frac{1+\eta}{\psi}\right) y_t - \frac{1+\eta}{\psi} z_t - \epsilon_{Ct} + \epsilon_{Ht}$$

While the aggregate TFP (3.14) linearises as:

$$-(1+\eta) z_t = \int -(1+\eta)(a_t(j) + v_{y,t}(j)) - \varepsilon (1+\eta)(p_t(j) - p_t)\, dj$$

But since $p_t = \int p_t(j)\, dj$ in a linear approximation and $\int v_{y,t}(j)\, dj = 0$ by definition, this is just:

$$z_t = \int a_t(j)\, dj = \epsilon_{At}$$

so that the equilibrium real wage in period t is given by:

$$w_t - p_t = \left(\frac{1}{\sigma} + \frac{1+\eta}{\psi}\right) y_t - \frac{1+\eta}{\psi} \epsilon_{At} - \epsilon_{Ct} + \epsilon_{Ht} \tag{3.33}$$

For reference, recall that σ is the elasticity of intertemporal substitution, ψ is the Frisch elasticity of labour supply and η is the elasticity of marginal cost.

3.A.3 Firms' marginal costs

Linearising the firm's marginal cost (3.9) and demand (3.5) gives:

$$mc_t(j) = w_t(j) - p_t + \eta y_t(j) - (1+\eta) a_t(j) \tag{3.34}$$

$$y_t(j) = y_t + v_{y,t}(j) - \varepsilon (p_t(j) - p_t) \tag{3.35}$$

Substituting the latter of these into the former gives:

$$mc_t(j) = (w_t(j) - p_t) + \eta y_t + \eta v_{y,t}(j) - \eta\varepsilon (p_t(j) - p_t) - (1+\eta) a_t(j)$$

Substituting in (3.32) and (3.33) for j's real wage, we then obtain:

$$mc_t(j) = \left(\eta + \frac{1}{\sigma} + \frac{1+\eta}{\psi}\right) y_t + \eta v_{y,t}(j) - \eta\varepsilon (p_t(j) - p_t) + \omega_t(j) \tag{3.36}$$

where $\omega_t(j)$ is the combined supply shock for firm j in period t, defined as:

$$\omega_t(j) \equiv B\boldsymbol{x}_t + Q\boldsymbol{v}_t(j) \tag{3.37a}$$

$$B = \begin{bmatrix} -(1+\eta)\left(1+\frac{1}{\psi}\right) & -1 & 1 & 0 \end{bmatrix} \tag{3.37b}$$

$$Q = \begin{bmatrix} -(1+\eta) & 1 & 0 \end{bmatrix} \tag{3.37c}$$

Next, we define $\dddot{mc}_t(j)$ to be a *partial average* of the firm's real marginal cost: the real marginal cost that firm j would incur without idiosyncratic demand shocks and if called upon to produce the average quantity (i.e. if $v_{y,t}(j) = 0$ and $y_t(j) = y_t$), but still with idiosyncratic supply shocks:

$$\begin{aligned} \dddot{mc}_t(j) &\equiv (w_t(j) - p_t) + \eta y_t - (1+\eta)\, a_t(j) \\ &= \left(\eta + \frac{1}{\sigma} + \frac{1+\eta}{\psi}\right) y_t + \omega_t(j) \end{aligned} \tag{3.38}$$

Finally, we define $\overline{mc}_t$ as the (true) average real marginal cost. That is, the real marginal cost a firm would incur if facing the average demand and experiencing the average supply shock (i.e. if producing the average quantity of output *and* experiencing no idiosyncratic shocks):

$$\begin{aligned} \overline{mc}_t &\equiv (w_t - p_t) + \eta y_t - (1+\eta)\, \epsilon_{At} \\ &= \left(\eta + \frac{1}{\sigma} + \frac{1+\eta}{\psi}\right) y_t + B\boldsymbol{x}_t \end{aligned} \tag{3.39}$$

3.A.4 Firms' price-setting under static pricing

When all firms are free to adjust their prices every period, prices will be expressed as a simple markup over their expected nominal marginal costs:

$$P_t(j) = \left(\frac{\varepsilon}{\varepsilon - 1}\right) E_t(j)\left[P_t MC_t(j)\right]$$

Linearising this then gives the simple:

$$p_t(j) = E_t(j)\left[p_t + mc_t(j)\right]$$

Substituting in equation (3.36) and gathering like terms then gives:

$$\begin{aligned} p_t(j) &= E_t(j)\left[p_t + \frac{1}{1+\eta\epsilon}\left(\dddot{mc}_t(j) + \eta v_{y,t}(j)\right)\right] \\ &= E_t(j)\left[p_t + \frac{1}{1+\epsilon\eta}\dddot{mc}_t(j)\right] \end{aligned} \tag{3.40}$$

where $\dddot{mc}_t(j)$ is defined in equation (3.38) above and the second equality makes use of the fact that $v_{y,t}(j)$ is an entirely transitory shock and firms do not discover their demand until after setting their prices.

We next define $\chi \equiv \left(\frac{1}{1+\varepsilon\eta}\right)\left(\eta + \frac{1}{\sigma} + \frac{1+\eta}{\psi}\right)$ and obtain the aggregate price level by taking the simple average of (3.40):

$$p_t = \overline{E}_t\left[p_t + \chi y_t\right] + \left(\frac{1}{1+\varepsilon\eta}\right)\int_0^1 E_t(j)\left[\omega_t(j)\right]$$

With our assumption that firms always observe $\omega_t(j)$ directly, this becomes:

$$\begin{aligned} p_t &= \overline{E}_t\left[p_t + \chi\left(\frac{1}{1+\sigma\kappa_y}E_t\left[y_{t+1} + \sigma(1-\kappa_\pi)(p_{t+1} - p_t)\right] + \boldsymbol{\mu}_y'\boldsymbol{x}_t\right)\right] + \left(\frac{1}{1+\varepsilon\eta}\right)B\boldsymbol{x}_t \\ &= \overline{E}_t\left[\chi\left(\frac{1}{1+\sigma\kappa_y}\left(y_{t+1} + \sigma(1-\kappa_\pi)p_{t+1}\right) + \boldsymbol{\mu}_y'\boldsymbol{x}_t\right)\right] + \left(\frac{1}{1+\varepsilon\eta}\right)B\boldsymbol{x}_t + \xi\overline{E}_t\left[p_t\right] \end{aligned}$$

where

$$\xi \equiv 1 - \left(\frac{1}{1+\varepsilon\eta}\right)\left(\eta + \frac{1}{\sigma} + \frac{1+\eta}{\psi}\right)\left(\frac{1}{1+\sigma\kappa_y}\right)\sigma(\kappa_\pi - 1)$$

Substituting this back into itself then eventually yields:

$$p_t = \sum_{k=0}^{\infty} \xi^k \overline{E}_t^{(k)} \left[\left(\frac{1}{1+\varepsilon\eta} \right) B\boldsymbol{x}_t + \overline{E}_t \left[\chi \left(\frac{1}{1+\sigma\kappa_y} \left(y_{t+1} + \sigma \left(1-\kappa_\pi\right) p_{t+1} \right) + \boldsymbol{\mu}_y' \boldsymbol{x}_t \right) \right] \right]$$
$$= \left(\frac{1}{1+\varepsilon\eta} \right) B\boldsymbol{x}_t + \sum_{k=0}^{\infty} \xi^k \overline{E}_t^{(k+1)} \begin{bmatrix} \left(\chi \boldsymbol{\mu}_y' + \xi \left(\frac{1}{1+\varepsilon\eta} \right) B \right) \boldsymbol{x}_t \\ +\chi \left(\frac{1}{1+\sigma\kappa_y} \right) \left(y_{t+1} + \sigma \left(1-\kappa_\pi\right) p_{t+1} \right) \end{bmatrix}$$

which is equation (3.29a) in the main text.

3.A.5 Solving the model under static pricing, part 1: Coefficients for aggregate variables

We have the following conjectured solution to the model:

$$\begin{aligned} X_t &= FX_{t-1} + G_1\boldsymbol{u}_t + G_2\boldsymbol{z}_t + G_3\boldsymbol{z}_{t-1} + G_4\boldsymbol{e}_t \\ p_t &= \boldsymbol{\gamma}_p' X_t \\ y_t &= \boldsymbol{\gamma}_y' X_t + \boldsymbol{\delta}_y'\boldsymbol{z}_t \\ w_t - p_t &= \boldsymbol{\gamma}_\varpi' X_t + \boldsymbol{\delta}_\varpi'\boldsymbol{z}_t \\ h_t &= \boldsymbol{\gamma}_h' X_t + \boldsymbol{\delta}_h'\boldsymbol{z}_t \end{aligned}$$

and here confirm this structure by deriving expressions for the $\boldsymbol{\gamma}_*$ and $\boldsymbol{\delta}_*$ coefficients. Firms' signal extraction problem and the law of motion for X_t are addressed below in section 3.A.6.

Real GDP

Starting with the linearised expression for aggregate demand (3.27) and making use of the conjectured solution, we have:

$$\begin{aligned} y_t &= \frac{1}{1+\sigma\kappa_y} E_t\left[y_{t+1} + \sigma\left(1-\kappa_\pi\right)\left(p_{t+1}-p_t\right)\right] + \boldsymbol{\mu}_y'\boldsymbol{x}_t \\ &= \frac{1}{1+\sigma\kappa_y} E_t\left[\gamma_y' X_{t+1} + \sigma\left(1-\kappa_\pi\right)\left(\gamma_p' X_{t+1} - \gamma_p' X_t\right)\right] + \boldsymbol{\mu}_y'\boldsymbol{x}_t \end{aligned}$$

We next note that since the household and central bank have full information, their expectation of the next-period state will be given by:

$$E_t\left[X_{t+1}\right] = FX_t + G_3\boldsymbol{z}_t$$

Making use of this, we can then write:

$$\begin{aligned} y_t &= \frac{1}{1+\sigma\kappa_y}\left(\gamma_y'\left(FX_t + G_3\boldsymbol{z}_t\right) + \sigma\left(1-\kappa_\pi\right)\left(\gamma_p'\left(FX_t + G_3\boldsymbol{z}_t\right) - \gamma_p' X_t\right)\right) + \boldsymbol{\mu}_y'\boldsymbol{x}_t \\ &= \left\{\boldsymbol{\mu}_y' S_x + \frac{1}{1+\sigma\kappa_y}\left(\gamma_y' F + \sigma\left(1-\kappa_\pi\right)\gamma_p'\left(F-I\right)\right)\right\} X_t \\ &+ \left\{\frac{1}{1+\sigma\kappa_y}\left(\gamma_y' + \sigma\left(1-\kappa_\pi\right)\gamma_p'\right) G_3\right\}\boldsymbol{z}_t \end{aligned}$$

which is to say that

$$\boldsymbol{\gamma}_y' = \boldsymbol{\mu}_y' S_x + \frac{1}{1+\sigma\kappa_y}\left(\boldsymbol{\gamma}_y' F + \sigma\left(1-\kappa_\pi\right)\boldsymbol{\gamma}_p'\left(F-I\right)\right)$$

$$\boldsymbol{\delta}_y' = \frac{1}{1+\sigma\kappa_y}\left(\boldsymbol{\gamma}_y' + \sigma\left(1-\kappa_\pi\right)\boldsymbol{\gamma}_p'\right)G_3$$

Gathering the terms in $\boldsymbol{\gamma}_y$ then gives

$$\boldsymbol{\gamma}_y' = \left(\boldsymbol{\mu}_y' S_x + \frac{\sigma}{1+\sigma\kappa_y}\left(1-\kappa_\pi\right)\boldsymbol{\gamma}_p'\left(F-I\right)\right)\left(I - \frac{1}{1+\sigma\kappa_y}F\right)^{-1}$$

Hours and the real wage

Starting with the expression for the equilibrium real wage (3.33) and substituting in the conjectured solution, we have

$$w_t - p_t = \left(\frac{1}{\sigma} + \frac{1+\eta}{\psi}\right)\left(\boldsymbol{\gamma}_y' X_t + \boldsymbol{\delta}_y' \boldsymbol{z}_t\right) + \begin{bmatrix} -\frac{1+\eta}{\psi} & -1 & 1 & 0 \end{bmatrix}\boldsymbol{x}_t$$

or, gathering terms,

$$\begin{aligned} w_t - p_t = &\left\{\left(\frac{1}{\sigma} + \frac{1+\eta}{\psi}\right)\boldsymbol{\gamma}_y' + \begin{bmatrix} -\frac{1+\eta}{\psi} & -1 & 1 & 0 \end{bmatrix} S_x\right\} X_t \\ &+ \left(\frac{1}{\sigma} + \frac{1+\eta}{\psi}\right)\boldsymbol{\delta}_y' \boldsymbol{z}_t \end{aligned}$$

from which we can immediately read that

$$\boldsymbol{\gamma}_\varpi' = \left(\frac{1}{\sigma} + \frac{1+\eta}{\psi}\right)\boldsymbol{\gamma}_y' + \begin{bmatrix} -\frac{1+\eta}{\psi} & -1 & 1 & 0 \end{bmatrix} S_x$$

$$\boldsymbol{\delta}_\varpi' = \left(\frac{1}{\sigma} + \frac{1+\eta}{\psi}\right)\boldsymbol{\delta}_y'$$

Linearising the aggregate production function (3.13) and making use of the fact that $z_t = \epsilon_{At}$ (shown above) then gives us

$$\begin{aligned} h_t &= \left(1+\eta\right)\left(y_t - \epsilon_{At}\right) \\ &= \left(1+\eta\right)\left(\boldsymbol{\gamma}_y' X_t + \boldsymbol{\delta}_y' \boldsymbol{z}_t\right) - \left(1+\eta\right)\begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}\boldsymbol{x}_t \end{aligned}$$

so that

$$\boldsymbol{\gamma}_h' = \left(1+\eta\right)\boldsymbol{\gamma}_y' - \left(1+\eta\right)\begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} S_x$$

$$\boldsymbol{\delta}_h' = \left(1+\eta\right)\boldsymbol{\delta}_y'$$

The aggregate price level

Substituting the conjectured solution into the expression for the aggregate price level (3.29a), we obtain

$$p_t = BS_xX_t + \sum_{k=0}^{\infty} \xi^k \overline{E}_t^{(k+1)} \begin{bmatrix} \left(\chi \boldsymbol{\mu}_y' + \xi B\right) S_x X_t \\ +\chi \left(\frac{1}{1+\sigma\kappa_y}\right) \left(\boldsymbol{\gamma}_y X_{t+1} + \boldsymbol{\delta}_y \boldsymbol{z}_{t+1}\right) \\ +\chi \left(\frac{1}{1+\sigma\kappa_y}\right) \sigma \left(1-\kappa_\pi\right) \boldsymbol{\gamma}_p' X_{t+1} \end{bmatrix}$$

Next, noting that firms have no knowledge of the current-period or future network shocks

$$E_t\left(j\right)\left[\boldsymbol{z}_{t+s}\right] = \mathbf{0} \; \forall s \geq 0$$

we consequently have that

$$E_t\left(j\right)\left[X_{t+1}\right] = FE_t\left(j\right)\left[X_t\right] \;\; \forall j$$

and, as such, the aggregate price level is given by

$$\begin{aligned} p_t &= BS_xX_t + \sum_{k=0}^{\infty} \xi^k \overline{E}_t^{(k+1)} \begin{bmatrix} \left(\chi \boldsymbol{\mu}_y' + \xi B\right) S_x X_t \\ +\chi \left(\frac{1}{1+\sigma\kappa_y}\right) \boldsymbol{\gamma}_y' F X_t \\ +\chi \left(\frac{1}{1+\sigma\kappa_y}\right) \sigma \left(1-\kappa_\pi\right) \boldsymbol{\gamma}_p' F X_t \end{bmatrix} \\ &= BS_xX_t + \begin{pmatrix} \left(\chi \boldsymbol{\mu}_y' + \xi B\right) S_x \\ +\chi \left(\frac{1}{1+\sigma\kappa_y}\right) \boldsymbol{\gamma}_y' F \\ +\chi \left(\frac{1}{1+\sigma\kappa_y}\right) \sigma \left(1-\kappa_\pi\right) \boldsymbol{\gamma}_p' F \end{pmatrix} \sum_{k=0}^{\infty} \xi^k \overline{E}_t^{(k+1)} \left[X_t\right] \end{aligned}$$

Noting that $\overline{E}^{(k)}\left[X_t\right] = T_s^k X_t$, this becomes

$$p_t = \left\{ BS_x + \begin{pmatrix} \left(\chi \boldsymbol{\mu}_y' + \xi B\right) S_x \\ +\chi \left(\frac{1}{1+\sigma\kappa_y}\right) \boldsymbol{\gamma}_y' F \\ +\chi \left(\frac{1}{1+\sigma\kappa_y}\right) \sigma \left(1-\kappa_\pi\right) \boldsymbol{\gamma}_p' F \end{pmatrix} \left(I - \xi T_s\right)^{-1} T_s \right\} X_t$$

from which we can immediately read that

$$\boldsymbol{\gamma}_p' = BS_x + \begin{pmatrix} \left(\chi \boldsymbol{\mu}_y' + \xi B\right) S_x \\ +\chi \left(\frac{1}{1+\sigma\kappa_y}\right) \boldsymbol{\gamma}_y' F \\ +\chi \left(\frac{1}{1+\sigma\kappa_y}\right) \sigma \left(1-\kappa_\pi\right) \boldsymbol{\gamma}_p' F \end{pmatrix} \left(I - \xi T_s\right)^{-1} T_s$$

Finally, we gather the terms in $\boldsymbol{\gamma}_p'$ to arrive at

$$\boldsymbol{\gamma}_p' = \left(BS_x + \begin{pmatrix} \left(\chi\boldsymbol{\mu}_y' + \xi B\right) S_x \\ +\chi\left(\frac{1}{1+\sigma\kappa_y}\right)\boldsymbol{\gamma}_y' F \end{pmatrix} \left(I - \xi T_s\right)^{-1} T_s \right)$$
$$\times \left(I - \left(1-\xi\right) F \left(I - \xi T_s\right)^{-1} T_s\right)^{-1}$$

3.A.6 Solving the model under static pricing, part 2: Firms' learning and the evolution of X_t

In order to characterise the law of motion for the hierarchy of firms' expectations, we need to first derive expressions for the signals they receive. We here step through these in order before moving onto the firms' signal extraction problem.

The (current period) combined supply shock

From equation (3.37), firm j's combined supply shock is given by:

$$\begin{aligned} \omega_t\left(j\right) &\equiv B\boldsymbol{x}_t + Q\boldsymbol{v}_t\left(j\right) \\ B &= \begin{bmatrix} -\left(1+\eta\right)\left(1+\frac{1}{\psi}\right) & -1 & 1 & 0 \end{bmatrix} \\ Q &= \begin{bmatrix} -\left(1+\eta\right) & 1 & 0 \end{bmatrix} \end{aligned}$$

where

$$\boldsymbol{v}_t\left(j\right) = \begin{bmatrix} v_{a,t}\left(j\right) \\ v_{w,t}\left(j\right) \\ v_{y,t}\left(j\right) \end{bmatrix}$$

The (previous period) quantity demanded

Recall that firm j's linearised demand function (3.35) is given by:

$$y_t\left(j\right) = y_t + v_{y,t}\left(j\right) - \varepsilon\left(p_t\left(j\right) - p_t\right)$$

Since firm j must have known their own price with certainty, news from the previous period's quantity demanded must come in the form:

$$y_{t-1}\left(j\right) + \varepsilon\, p_{t-1}\left(j\right) = y_{t-1} + v_{y,t-1}\left(j\right) + \varepsilon\, p_{t-1}$$

Making use of the posited solution and gathering like terms then gives:

$$y_{t-1}\left(j\right) + \varepsilon\, p_{t-1}\left(j\right) = \left(\boldsymbol{\gamma}_y' + \varepsilon\,\boldsymbol{\gamma}_p'\right) X_{t-1} + \boldsymbol{\delta}_y'\boldsymbol{z}_{t-1} + \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \boldsymbol{v}_{t-1}\left(j\right)$$

The (previous period) prices set by individual competitors

From equation (3.40), note that firm j's pricing decision is given by:

$$p_t(j) = E_t(j)\left[p_t + \frac{1}{1+\epsilon\eta}\dddot{mc}_t(j)\right]$$

and their partial average marginal cost (3.38) by:

$$\dddot{mc}_t(j) = \left(\eta + \frac{1}{\sigma} + \frac{1+\eta}{\psi}\right) y_t + \omega_t(j)$$

Making use of the posited solution and gathering like terms, we can therefore write the pricing rule as:

$$p_t(j) = \boldsymbol{\lambda}_1'\boldsymbol{x}_t + \boldsymbol{\lambda}_2' E_t(j)[X_t] + \boldsymbol{\lambda}_3'\boldsymbol{v}_{t-1}(j) \tag{3.41}$$

where

$$\boldsymbol{\lambda}_1 = \frac{1}{1+\varepsilon\eta}B' \tag{3.42a}$$

$$\boldsymbol{\lambda}_2 = \boldsymbol{\gamma}_p + \chi\boldsymbol{\gamma}_y \tag{3.42b}$$

$$\boldsymbol{\lambda}_3 = \frac{1}{1+\varepsilon\eta}Q' \tag{3.42c}$$

Recall that it will be necessary to step this back one period in order to consider firm i in period t observing $g_{t-1}(j)$ where $j = \delta_{t-1}(i)$.

If we take the simple average of equation (3.41), we get:

$$p_t = \underbrace{(\boldsymbol{\lambda}_1' S_x + \boldsymbol{\lambda}_2' T_s)}_{=\gamma_p'} X_t$$

This represents an alternative way of deriving the $\boldsymbol{\gamma}_p$ vector and so is a handy way of confirming the logic of the previous section.

Firms' signal extraction problem

Given the posited solution, we therefore have that agents observe the following private signal:

$$\boldsymbol{s}_t^p(i) = \begin{bmatrix} \omega_t(i) \\ y_{t-1}(i) + \varepsilon\, p_{t-1}(i) \end{bmatrix} = D_1\boldsymbol{x}_t + D_2 X_{t-1} + R_1\boldsymbol{v}_t(i) + R_2\boldsymbol{z}_{t-1}$$

where

$$D_1 = \begin{bmatrix} B \\ \mathbf{0} \end{bmatrix} \quad D_2 = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\gamma}_y' + \varepsilon \boldsymbol{\gamma}_p' \end{bmatrix} \quad R_1 = \begin{bmatrix} Q \\ \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \end{bmatrix} \quad R_2 = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\delta}_y' \end{bmatrix}$$

These, and the expression for individual firms' prices (3.41), are in format used in theorem 1 of chapter 2 and since the model here also satisfies assumptions 1 and 2 of the same chapter, theorem 1 therefore holds. In aggregate, the hierarchy of expectations will therefore follow the law of motion:

$$X_t \equiv \begin{bmatrix} \boldsymbol{x}_t \\ \overline{E}_t \left[X_t\right] \\ \widetilde{E}_t \left[X_t\right] \\ \widehat{E}_t \left[X_t\right] \end{bmatrix} = F X_{t-1} + G_1 \boldsymbol{u}_t + G_2 \boldsymbol{z}_t + G_3 \boldsymbol{z}_{t-1}$$

Appendix 3.B An irregular network is a stable equilibrium

In this appendix we demonstrate a first step towards proving the conjecture that an irregular observation network among price-setting firms is a stable equilibrium. Looking at a simplified setting where a continuum of firms each decides on an *ex-ante* probability with which they will observe each of just two candidates, (A)nne or (B)ill, we show that symmetric mixed strategies (i.e. everyone assigning the same probability to A) will be an equilibrium. To begin, we first define

$$\alpha_i \equiv \Pr\left(\text{Agent } i \text{ observes } A\right)$$

Given a common payoff function, $\pi\left(\alpha_i, \alpha_{-i}\right)$, where α_{-i} represents the α's of all agents except agent i, we wish to consider the problem in which agents solve

$$\max_{\alpha_i} \pi\left(\alpha_i, \alpha_{-i}\right) \text{ subject to } \alpha_{-i} = \overline{\alpha}$$

for some value $\overline{\alpha} \in (0, 1)$. It'll therefore be an equilibrium if

$$\pi_1\left(\overline{\alpha}, \overline{\alpha}\right) = 0$$

So, what is $\pi\left(\alpha_i, \alpha_{-i}\right)$? We take a very simple example with just static pricing. The optimal price for firm i is

$$p_t^*\left(i\right) = p_t + \chi y_t$$

where p is the average price, y is real GDP and χ is an inverse measure of strategic complementarity. The economy is cash-in-advance so that

$$m_t = y_t + p_t$$

and so

$$p_t^*\left(i\right) = \left(1 - \chi\right) p_t + \chi m_t$$

Note that $\chi = 0$ corresponds to complete strategic complementarity, while $\chi = 1$ corresponds to no strategic complementarity at all. Not observing either of the elements on the right hand side, agent i instead chooses to minimise the expected square of their deviation from this optimum

$$-\frac{1}{2}E_t(i)\left[(p_t(i)-p_t^*(i))^2\right]$$

and so sets

$$p_t(i)=E_t(i)\left[p_t^*(i)\right]$$

Selecting α_i then amounts to choosing what value of α_i will provide the best estimate of $p_t^*(i)$ in the sense of minimising the mean (i.e. expected) square error.

$$\pi(\alpha_i,\overline{\alpha})=-\frac{1}{2}E\left[(E_t(i)\left[p_t^*(i)\right]-p_t^*(i))^2\right]$$

Substituting in our expression for i's optimal price gives

$$\pi(\alpha_i,\overline{\alpha})=-\frac{1}{2}E\left[((1-\chi)\left\{E_t(i)\left[p_t\right]-p_t\right\}+\chi\left\{E_t(i)\left[m_t\right]-m_t\right\})^2\right]$$

Since we have a continuum of agents, agent i's contribution to the average price can be ignored. Suppose that the price chosen by firm j when observing A is given by

$$p_t(j,A)=\delta p_t(A)+e_t(j)$$

where $p(A)$ is the price set by firm A and $E\left[e_t(j)\right]=0\ \forall j,t$. This last requirement will hold when both firm j and firm A receive unbiased and independent signals regarding m_t and m_t has an unconditional expectation of zero. The equivalent setting applies for when j observes B. Then the average price will be given by

$$p_t=\overline{\alpha}\delta p_t(A)+(1-\overline{\alpha})\,\delta p_t(B)$$

Now consider agent i's expectation conditional on observing A:

$$\begin{aligned}E\left[p_t|p_t(A)\right]&=\overline{\alpha}\delta p_t(A)+(1-\overline{\alpha})\,\delta E\left[p_t(B)\,|p_t(A)\right]\\&=\overline{\alpha}\delta p_t(A)+(1-\overline{\alpha})\,\delta p_t(A)\\&=\delta p_t(A)\end{aligned}$$

where the second equality assumes that the prices of A and B are unbiased signals of each other. Agent i's payoff is then

$$\pi(\alpha_i,\overline{\alpha})=-\frac{1}{2}E\left[\left(\begin{array}{l}(1-\chi)\left\{\begin{array}{l}\left[\alpha_i(\delta p_t(A)+e_t(i))+(1-\alpha_i)(\delta p_t(B)+e_t(i))\right]\\-\delta\left[\overline{\alpha}p_t(A)+(1-\overline{\alpha})\,p_t(B)\right]\end{array}\right\}\\+\chi\left\{E_t(i)\left[m_t\right]-m_t\right\}\end{array}\right)^2\right]$$

Supposing still further that an observation of A and an observation of B are equally useful in improving agent i's estimate of m_t, we can then see that agent i's first order condition in their selection of α_i is given by

$$-E\left[\begin{array}{l}(1-\chi)\left\{\begin{array}{l}\left[\alpha_i\left(\delta p_t\left(A\right)+e_t\left(i\right)\right)+\left(1-\alpha_i\right)\left(\delta p_t\left(B\right)+e_t\left(i\right)\right)\right] \\ -\delta\left[\overline{\alpha}p_t\left(A\right)+\left(1-\overline{\alpha}\right)p_t\left(B\right)\right]\end{array}\right\} \\ +\chi\left\{E_t\left(i\right)\left[m_t\right]-m_t\right\}\end{array}\right]\left(\delta p_t\left(A\right)-\delta p_t\left(B\right)\right)=0$$

which simplifies down to

$$E\left[\begin{array}{l}(1-\chi)\left\{\begin{array}{l}\alpha_i\left(\delta p_t\left(A\right)+e_t\left(i\right)\right) \\ +\left(1-\alpha_i\right)\left(\delta p_t\left(B\right)+e_t\left(i\right)\right)-\overline{\alpha}\delta p_t\left(A\right)-\left(1-\overline{\alpha}\right)\delta p_t\left(B\right)\end{array}\right\} \\ +\chi\left\{E_t\left(i\right)\left[m_t\right]-m_t\right\}\end{array}\right]=0$$

On average, agent i's expectation of m_t will be correct $\left(E\left[E_t\left(i\right)\left[m_t\right]\right]=m_t\right)$ so this just becomes

$$E\left[\left\{\alpha_i\left(\delta p_t\left(A\right)+e_t\left(i\right)\right)+\left(1-\alpha_i\right)\left(\delta p_t\left(B\right)+e_t\left(i\right)\right)-\overline{\alpha}p_t\left(A\right)-\left(1-\overline{\alpha}\right)p_t\left(B\right)\right\}\right]=0$$

Since the unconditional expectation of $e_t\left(i\right)$ is zero, this collapses to

$$\alpha_i=\overline{\alpha}$$

as required!

Chapter 4

Price-setting under asymmetric TransLog preferences and incomplete information

Abstract

I explore firms' optimal price-setting behaviour when facing TransLog household preferences. I first solve explicitly for a firm's best-response pricing rule under full information, including an endogenous market-exit condition, and next show that in partial equilibrium under incomplete information, larger firms will focus more on movements in marginal cost while smaller firms will place more weight on changes in consumer preferences and competitors' prices. In general equilibrium, I characterise and estimate the effect of two distinct sources of real rigidity that emerge from TransLog preferences: first, the well-known curvature in demand and, second, the dramatic increase in complexity of firms' signal-extraction problems. Because household preferences are not fully uniform, the model also represents a channel through which firms' transitory idiosyncratic shocks can result in persistent aggregate volatility.

4.1 Introduction

Many – indeed, the vast majority of – macroeconomic models that employ monopolistic competition make use of the Dixit and Stiglitz (1977) aggregator for a representative household's preferences across individual consumption goods. This choice is motivated by both the analytical ease with which it is deployed and the paucity of controlling parameters, which aids in estimation. However, the Dixit-Stiglitz aggregator produces demand functions with constant (and, for the simplest and most common case, common) elasticities of demand for every good and, consequently, constant optimal mark-ups over marginal costs in firms' price-setting decisions.

This then imposes that each firm's consideration of other firms' prices comes *only* through consideration of its own nominal marginal costs. Strategic complementarity is therefore limited to only emerging through nominal input prices (such as wages) and the effect of aggregate demand on real marginal costs. True Bertrand competition is effectively assumed away.

That firms' mark-ups vary over time is a well-established fact, however. Standard practice in the literature has therefore been to suppose that mark-ups are subject to exogenous and persistent shocks,[1] an approach that seems odd given the explicit assumption (via the choice of the Dixit-Stiglitz aggregator) that they are constant.

This paper illustrates that by adopting a more realistic model of household demand, two important sources of real rigidity in price adjustment emerge, even when firms possess full and costless flexibility in their price setting. First, as initially explored by Kimball (1995), allowing firms' demand schedules to be curved (in log-log space) ensures that even when firms possess full information, an increase in their marginal costs will not be fully passed through to their prices because of consideration for the concomitant loss of demand. In other words, each firm's price setting rule involves taking an opinion on both its likely marginal cost and its optimal mark-up. Second, the inclusion of strategic complementarity through true price competition (i.e. the need to select its mark-up) poses each firm a dramatically more complex signal extraction problem when operating under incomplete information.

In particular, we here look at optimal price-setting for monopolistically competitive firms facing non-uniform TransLog preferences (a special case of the Almost Ideal Demand setting of Deaton and Muellbauer, 1980). Work by Bergin and Feenstra (2000) has previously looked at TransLog preferences as a means of achieving endogenous persistence following aggregate shocks. By working with a specific, parameterised model, they are able to avoid the Chari, Kehoe, and McGrattan (2000)

[1] See, for example, Smets and Wouters (2003).

criticism that the generalised Kimball (1995) aggregator permits arbitrarily strong curvature in demand. However, Bergin and Feenstra limit their attention to full uniformity in household preferences across goods and still require nominal rigidity[2] to achieve any persistence following a monetary shock. Given the recent challenges to nominal rigidity in individual prices, both from theoretical and empirical grounds (see chapter 1 for more detail), there is a need to explore sources of real rigidity that do not rely on individual price stickiness.

In contrast, the current chapter permits non-uniformity in household preferences across goods. We first consider firms' price-setting problem under full information. We derive an explicit, non-linear expression for each firm's price as a function of its marginal cost and the prices of its competitors[3] and demonstrate the existence of a unique Nash equilibrium in prices and an endogenous market exit condition.

In a linearised, partial equilibrium setting under uncertainty, we demonstrate (a) that in addition to needing to form an expectation of the aggregate price level in order to estimate its marginal cost, each firm must also estimate a *firm-specific* weighted-average of its competitors' prices in order to account for price competition; and (b) that larger firms will place relatively more weight on their marginal cost, while smaller firms will focus primarily on their competitors' prices and transitory shifts in the distribution of consumer demand.

In general equilibrium under uncertainty, full non-uniformity in preferences is unfortunately intractable, so we instead impose a setting of *near-uniformity in steady-state preferences*, wherein firms' steady-state mark-ups and shares of household expenditure are the same, but their prices and marginal costs are not. In other words, we allow for the existence of low-price, high-volume businesses alongside high-price, low-volume businesses in steady state.

In this setting, despite firms having full price-setting flexibility and access to public signals of aggregate variables, significant persistence in aggregate variables

[2]They suppose that firms operate under staggered contracts, with each firm's price fixed for two periods and half of all firms able to adjust in each period.

[3]Note that while a firm's price still equals a mark-up over its marginal cost, this only *implicitly* identifies its price under systems with curved demand since a price change will also affect its market share and, hence, its optimal mark-up.

emerges following shocks to both productivity and monetary policy. Finally, because of the non-uniformity in household preferences, the model presents a mechanism through which transitory idiosyncratic shocks cause persistent movements in aggregate variables. Using a parameterisation based on the distribution of firm sizes in the United States, we tentatively estimate that as much as 5% of observed volatility in the aggregate price level may be attributable to idiosyncratic factors.

This work therefore adds to existing literature of Gabaix (2011), Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Saleh (2012) and chapter 3 of this thesis in identifying an idiosyncratic source of aggregate volatility.

Methodologically, this chapter (like chapter 2) adapts and extends the techniques developed by Nimark (2008, 2011b) for finding the solution of incomplete information problems with strategic interaction when agents observe lagged signals of aggregate variables. As with chapter 2, the model here requires that agents consider multiple compound expectations (a simple-average and a weighted-average), but *unlike* that chapter, there is no requirement here to truncate the number of expectations in the agents' state vector of interest. That is, we here emerge with exactly two aggregated expectations of interest.

The remainder of this chapter is structured as follows. Section 4.2 provides a brief overview of TransLog preferences and the Almost Ideal Demand System. Section 4.3 next outlines the broader model of household and central bank behaviour and discusses the information available to each agent. Section 4.4 solves the price-setting problem under full information and illustrates the magnitude by which curvature of the demand curve can affect price changes. Section 4.5 then considers the firms' problem under incomplete information, both in general for partial equilibrium and with near-uniformity in preferences for general equilibrium. Section 4.6 illustrates these general equilibrium results by presenting a series of simulations following both aggregate and idiosyncratic shocks before section 4.7 concludes.

4.2 TransLog preferences and the Almost Ideal Demand System

This section presents a brief overview of TransLog preferences and the Almost Ideal Demand System (AIDS). Readers already familiar with these systems of demand may skip immediately to the model in section 4.3 below.

The Almost Ideal Demand System was originally devised by Deaton and Muellbauer (1980) as an empirical tool to permit the estimation of a generalised system of demand. It is consistent with standard theory of consumer optimisation and, by satisfying the conditions laid out by Muellbauer (1975, 1976), it aggregates exactly and so admits a representative consumer. Transcendental Logarithm (TransLog) preferences were developed earlier by Christensen, Jorgenson, and Tau (1975) and are nested entirely within the Almost Ideal setting. We shall therefore present a brief overview of the Almost Ideal framework first before then discussing the additional restrictions required to obtain TransLog preferences.

4.2.1 The Almost Ideal Demand System

The Almost Ideal Demand System is itself based on the PIGLOG (price-independent, generalised, linear-in-logarithms) model of consumer preferences, in which individual preferences are described via the expenditure function:

$$\ln\left[e\left(u, \boldsymbol{P}\right)\right] = \left(1-u\right)\ln\left[a\left(\boldsymbol{P}\right)\right] + u\ln\left[b\left(\boldsymbol{P}\right)\right] \tag{4.1}$$

where $\boldsymbol{P}$ is a vector of all prices and, with some exceptions,[4] u varies from 0 (subsistence) to 1 (bliss). The AIDS model then proposes particular functional forms for $\ln\left[a\left(\boldsymbol{P}\right)\right]$ and $\ln\left[b\left(\boldsymbol{P}\right)\right]$, namely:

$$\ln\left[a\left(\boldsymbol{P}\right)\right] = \alpha_0 + \sum_i \alpha\left(i\right)\ln\left(P\left(i\right)\right) + \frac{1}{2}\sum_i\sum_j \gamma^*_{ij}\ln\left(P\left(i\right)\right)\ln\left(P\left(j\right)\right) \tag{4.2a}$$

$$\ln\left[b\left(\boldsymbol{P}\right)\right] = \ln\left[a\left(\boldsymbol{P}\right)\right] + \beta_0\prod_i P\left(i\right)^{\beta_i} \tag{4.2b}$$

[4] See the appendix of Deaton and Muellbauer (1980) for more detail.

so that the expenditure function (4.1) becomes:

$$\begin{aligned}\ln\left[e\left(u,\boldsymbol{P}\right)\right]=\alpha_0+\sum_i\alpha\left(i\right)\ln\left(P\left(i\right)\right)\\+\frac{1}{2}\sum_i\sum_j\gamma^*_{ij}\ln\left(P\left(i\right)\right)\ln\left(P\left(j\right)\right)+u\beta_0\prod_i P\left(i\right)^{\beta_i}\end{aligned}\tag{4.3}$$

Under this specification, the Marshallian (i.e. uncompensated) demand function for good i, expressed as a share of total nominal expenditure, is:

$$\frac{P\left(i\right)Q\left(i\right)}{N}\equiv s\left(i\right)=\alpha\left(i\right)+\sum_j\gamma_{ij}\ln\left(P\left(j\right)\right)+\beta_i\ln\left(\frac{N}{\mathbb{P}}\right)\tag{4.4}$$

where $\gamma_{ik}\equiv\frac{1}{2}\left(\gamma^*_{ik}+\gamma^*_{ki}\right)$, N is income (and total nominal expenditure) and $\mathbb{P}$ is the aggregate price index, defined as:

$$\ln\left(\mathbb{P}\right)\equiv\alpha_0+\sum_i\alpha\left(i\right)\ln\left(P\left(i\right)\right)+\frac{1}{2}\sum_i\sum_j\gamma_{ij}\ln\left(P\left(i\right)\right)\ln\left(P\left(j\right)\right)\tag{4.5}$$

The following restrictions are then added for the system to comply with standard consumer theory:

$$\sum_{i=1}^{J}\alpha\left(i\right)=1,\ \sum_{i=1}^{J}\gamma_{ij}=0\,\forall j\text{ and }\sum_{i=1}^{J}\beta_i=0\tag{4.6}$$

for adding up (i.e. to ensure that $\sum s\left(i\right)=1$);

$$\sum_{j=1}^{J}\gamma_{ij}=0\,\forall i\tag{4.7}$$

to ensure that demand functions are homogeneous of degree zero (and the expenditure function homogenous of degree one); and

$$\gamma^*_{ij}=\gamma^*_{ji}\,\forall i,j\tag{4.8}$$

to ensure the symmetry of the substitution matrix.

The following set of elasticities of demand may then be derived from equation (4.4). Assuming that total nominal expenditure (i.e. income) does not change with movements in individual prices and conditional on the prices of all other goods, the (positive) own-price elasticity of demand for good i is:

$$\begin{aligned}\varepsilon_{ii} &= \left|\frac{\partial \ln Q(i)}{\partial \ln P(i)}\right| \\ &= 1 - \frac{\partial \ln s(i)}{\partial \ln P(i)} \\ &= 1 - \frac{1}{s(i)}\left[\gamma_{ii} - \beta_i\left(\alpha(i) + \sum_k \gamma_{ki} \ln(P(k))\right)\right], \end{aligned} \tag{4.9}$$

the cross-price elasticity of demand for good i following a change in the price of good j is:

$$\varepsilon_{ij} = \frac{1}{s(i)}\left[\gamma_{ij} - \beta_i\left(\alpha(j) + \sum_k \gamma_{kj} \ln(P(k))\right)\right]; \text{ and} \tag{4.10}$$

the income elasticity of demand for good i is:

$$\eta_i = 1 + \frac{\beta_i}{s(i)} \tag{4.11}$$

It is immediately apparent that this framework possesses the potential for considerably richer dynamics following an aggregate shock than might be expected with the Dixit-Stiglitz or Kimball aggregators. With each firm's elasticity being dependent on a weighted sum of all other firms' prices, optimal mark-ups will be both time-varying and different for every firm. Indeed, the super-elasticity of demand – the elasticity of the own-price elasticity, sometimes called the curvature of demand – can be shown (a derivation is provided in appendix 4.A.1) to be:

$$\begin{aligned}\xi_{ii} &\equiv \frac{\partial \ln \varepsilon_{ii}}{\partial \ln P(i)} \\ &= \frac{1}{\varepsilon_{ii}}\left[(\varepsilon_{ii} - 1)^2 + \frac{\beta_i \gamma_{ii}}{s(i)}\right] \end{aligned} \tag{4.12}$$

4.2.2 TransLog preferences

TransLog preferences are nested within the Almost Ideal system. They are obtained by supposing that the income elasticity of demand is unitary for all goods (i.e. there are no luxury or necessary goods).

$$\beta_i = 0 \Leftrightarrow \eta_i = 1 \ \forall i \tag{4.13}$$

In other words, TransLog preferences are the AIDS model with the further imposition that preferences be *homothetic*.[5] This then reduces the expression for i's share of expenditure to

$$s(i) = \alpha(i) + \sum_j \gamma_{ij} \ln(P(j)) \tag{4.14}$$

and the expressions for cross-price elasticity, own-price elasticity and own-price super-elasticity of demand to

$$\varepsilon_{ij} = \frac{\gamma_{ij}}{s(i)}; \tag{4.15}$$

$$\varepsilon_{ii} = 1 - \frac{\gamma_{ii}}{s(i)}; \text{ and} \tag{4.16}$$

$$\xi_{ii} = \frac{(\varepsilon_{ii} - 1)^2}{\varepsilon_{ii}} \tag{4.17}$$

respectively. Note that in this setting, each good's super-elasticity is unambiguously positive and increasing in the own-price elasticity.

Supposing that good i is produced by a monopolist, the optimal mark-up over marginal costs for that good will be given by

$$\mu_i = \frac{\varepsilon_{ii}}{\varepsilon_{ii} - 1} = 1 - \frac{s(i)}{\gamma_{ii}} \tag{4.18}$$

so that we can rewrite the super-elasticity as

$$\xi_{ii} = \frac{\varepsilon_{ii} - 1}{\mu_i} = \frac{\varepsilon_{ii}}{\mu_i^2} \tag{4.19}$$

4.2.3 An initial comparison to other demand systems

In the near-ubiquitous CES demand system of Dixit and Stiglitz, where the own-price and cross-price elasticities of demand are common and constant, each firm will have a common and constant optimal mark-up over its (potentially different) marginal costs. In contrast, in the TransLog and Almost Ideal settings, when a firm raises its price in such a way as to lower its share of aggregate spending (i.e. in the absence

[5] Recall that a preference relation over bundles within $\mathbb{R}_+$ is homothetic if, when $x \sim y$, we also have that $\alpha x \sim \alpha y$ for any $\alpha \geq 0$.

of sufficient price increases from its competitors), its own-price elasticity of demand will rise, causing its optimal mark-up to fall and so partially offset the increase in price.

Furthermore, since the super-elasticity is strictly positive and increasing under TransLog preferences and generally so under AIDS,[6] this dampening effect becomes convexly stronger for larger price increases, thereby creating a strong incentive for firms to avoid lifting their prices above their aggregate reference prices.

In this respect, the AIDS framework is quite similar to the demand system implied by the Kimball (1995) aggregator. As with Kimball's preferences, aggregate price rigidity here will be highly sensitive to anything that impedes price coordination. In the absence of concrete knowledge that their competitors are also raising their prices, firms will temper any increases of their own, thereby increasing the persistence of the effects of any nominal shock to the economy. However, there are two key differences between the Kimball and Almost Ideal demand frameworks. First, where Kimball preferences allow for arbitrarily strong curvature in demand (a fact criticised as unrealistic by Chari, Kehoe, and McGrattan, 2000), the Almost Ideal system exhibits well-defined and moderate curvature in demand. Next, unlike the Kimball demand system, the Almost Ideal setting explicitly models asymmetries in household preferences across goods, both in their base market shares and their sensitivity to movements to other firms' prices. Because of the latter, each firm needs to consider a firm-specific reference price, formed as a weighted average of its competitors' prices, in addition to the aggregate price level.

It is, of course, possible to model unequal preferences across goods within the Dixit-Stiglitz setting by use of nested Constant Elasticity of Substitution (CES) functions. However, this approach is not readily able to be extended to unequal preferences over a continuum of goods and, in any event, will still produce constant elasticities of demand and so is unable to speak to the real rigidities embodied in time-varying mark-ups.

[6]Technically, the super-elasticity can be negative under AIDS for sufficiently large values of β_i, but empirical estimates of β_i are typically quite low. See also the discussion of Dossche, Heylen, and Van den Poel (2010) regarding the number of goods with positive super-elasticities.

While the assumption that firms' own-price elasticity can vary is a clear generalisation from the CES framework of Dixit-Stiglitz, the assumption here that the super-elasticity of demand is strictly positive is arguably still too restrictive. Dossche, Heylen, and Van den Poel (2010) examine scanner data for a large euro area retailer using a modified version of the Almost Ideal framework[7] and find that as many as 42% of goods have a negative super-elasticity (denoted "curvature" in their paper). However, the median super-elasticity across all items is positive (0.8) and higher still across non-food items (1.14). When limiting attention to items with an estimated (absolute value of) elasticity of unity or greater ($\varepsilon \geq 1$), only 26% of items have a negative super-elasticity and the median super-elasticity is 1.7. Elasticity and super-elasticity were found to be strongly positively correlated, with a correlation coefficient of 0.53.

[7] The authors add a behavioural extension to the AIDS model of Deaton and Muellbauer (1980), justified by an appeal to loss aversion, that permits them to freely estimate demand super-elasticities which would otherwise be fully determined by own-price elasticities.

4.3 The model

The model considered here is standard in its treatment of households' intertemporal decision making and labour supply; the central bank's monetary policy; and firms' production technologies. It differs principally from the basic New Keynesian model in three aspects: first, that TransLog preferences are used to capture differentiation across consumption goods; second, that those preferences are not assumed to be symmetric over goods; and third, that firms operate under incomplete information. In order to emphasise the real rigidity invoked by the model, prices are assumed to be perfectly flexible.

Each of these innovations has been separately explored in the literature. Bergin and Feenstra (2000) demonstrated endogenous persistence following monetary shocks with TransLog preferences, staggered pricing and a production structure that used the final good as an intermediate good, but also made the simplifying assumption of full uniformity in preferences. Woodford (2003) examined the persistence obtained from firms' possessing incomplete information regarding shocks to nominal GDP under Dixit-Stiglitz preferences and static (i.e. flexible) price setting. Nimark (2008) later extended this to include idiosyncratic shocks to marginal costs and dynamic pricing in the style of Calvo (1983). However, to our knowledge, these ideas have not previously been brought together in a single model.

Uncertainty will enter the model on both the supply and demand sides of the economy. On the supply side, firms will experience both aggregate and idiosyncratic shocks to their marginal costs in the form of movements in their productivity. On the demand side, aggregate shocks will be delivered by monetary policy, while idiosyncratic shocks will apply to households' relative preferences across goods.

Notation

We generally make use of the notation that an uppercase letter denotes the variable itself; a lowercase letter is the (natural) log of that variable; a variable with an asterisk denotes the value of that variable in steady-state (defined below); and a variable with a hat represent the deviation of that variable from its value in steady-state.[8]

$$x_t \equiv \ln(X_t) \qquad \widehat{x}_t \equiv x_t - x^*$$

The two exceptions to this rule will be firms' share of household expenditure ($s_t(i)$) and the interest rate (i_t). In both these cases, the lowercase letter denotes the variable itself. It remains the case that $\widehat{s}_t(i) = s_t(i) - s(i)^*$ and $\widehat{i}_t = i_t - i^*$.

4.3.1 The household

Each period, the representative household maximises[9]

$$E_t^{HH}\left[\sum_{s=0}^{\infty}\beta^s\left\{\frac{C_{t+s}^{1-\frac{1}{\sigma}}-1}{1-\frac{1}{\sigma}}-\frac{H_{t+s}^{1+\frac{1}{\psi}}}{1+\frac{1}{\psi}}\right\}\right] \tag{4.20}$$

subject to the budget constraint

$$W_tH_t + (1+i_{t-1})B_{t-1} + T_t = \mathbb{P}_tC_t + B_t \tag{4.21}$$

where $E_t^{HH}[\cdot]$ is the mathematical expectation conditional on the household's information set in period t (defined below); C_t is aggregate (real) consumption; H_t is the aggregate labour supply; σ is the elasticity of intertemporal substitution; ψ is the Frisch elasticity of labour supply; W_t is the nominal wage rate; B_t denotes holdings of one-period risk-free bonds; i_t is the nominal interest rate; and T_t combines firm profits and nominal lump sum transfers. The aggregate first-order conditions for the household's problem are therefore standard:

$$H_t^{\frac{1}{\psi}} = \frac{W_t}{\mathbb{P}_t}C_t^{-\frac{1}{\sigma}} \tag{4.22}$$

$$C_t^{-\frac{1}{\sigma}} = \beta(1+i_t)E_t^{HH}\left[C_{t+1}^{-\frac{1}{\sigma}}\frac{1}{\Pi_{t+1}}\right] \tag{4.23}$$

where $\Pi_t \equiv \mathbb{P}_t/\mathbb{P}_{t-1}$ is the gross rate of inflation.

[8] It should be noted that this notation differs slightly to that of the preceding chapters. This is necessary because the previous two chapters had no need to refer to the log of variables, only their log deviation, while this chapter refers to both.

[9] Note that β here is the household's discount factor, as distinct from the coefficient governing the income elasticity of demand used above.

Household preferences over differentiated consumption goods for a given amount of nominal expenditure ($N_t = \mathbb{P}_t C_t$) are represented in TransLog form, as described in section 4.2.2 above. We therefore have the following expression for household demand for good i:

$$C_t(i) = \frac{s_t(i)\,\mathbb{P}_t C_t}{P_t(i)} \tag{4.24}$$

In addition to those outlined earlier, we make two further assumptions regarding household preferences:

$$\alpha_t(i) = \frac{1}{J}\left[\zeta_i + v_t^{\alpha}(i)\right] \tag{4.25a}$$

$$\gamma_{ij} > 0 \;\forall j \neq i \tag{4.25b}$$

The first of these (4.25a) declares $\alpha_t(i)$, which we refer to as the firm's *base market share*,[10] to be the sum of an underlying, fundamental preference ($\alpha(i)^* \equiv \frac{1}{J}\zeta_i$) and a time-varying, stochastic component ($v_t^{\alpha}(i)$ is a mean zero shock, defined below). Recall that we must have $\sum_i \alpha_t(i) = 1$ in every period.

The second assumption (4.25b) states that all goods are gross substitutes for each other. This ensures that cross-price elasticities of demand are all strictly positive ($\varepsilon_{ij} > 0 \;\forall i \neq j$); that the absolute values of all own-price elasticities of demand are strictly greater than one ($\varepsilon_{ii} > 1 \;\forall i$);[11]; and that the vector of equilibrium prices under full information is unique (see section 4.4).

By comparison, existing literature on the use of TransLog preferences in macroeconomic models (see, for example, Bergin and Feenstra, 2000) has tended to suppose complete uniformity in preferences across goods by imposing the following restrictions on $\boldsymbol{\alpha}$ and Γ:

$$\alpha_t(i) = \frac{1}{J}\;\forall i, t \tag{4.26a}$$

$$\gamma_{ii} = -\frac{\gamma}{J}\;\forall i\;;\;\gamma_{ij} = \frac{\gamma}{J(J-1)}\;\forall j \neq i \tag{4.26b}$$

[10]Named such because, if all firms were to charge the same price, it would be their share of household expenditure

[11]Combined with the homogeneity restriction (4.7), (4.25b) ensures that that $\gamma_{ii} < 0 \;\forall i$ and, hence, that $\varepsilon_{ii} > 1$.

While easier to work with, these assumptions have shielded from view some important aspects of firms' price-setting behaviour and their informational requirements, as will be shown below in a partial equilibrium setting.

However, when moving to a general equilibrium setting, it will still be necessary to impose what we call *near-uniformity in steady-state preferences* in order to achieve tractability. This will include uniformity in Γ (the matrix of γ_{ij}, equation 4.26b) and uniformity in steady-state expenditure shares, but will retain the asymmetry in base market shares. Some of the dynamics implied by our partial equilibrium results will therefore not be present in our simulations. Section 4.3.7 covers this in more detail.

4.3.2 The firm

Production

Each good is produced by a single firm according to a common production function that deploys labour with decreasing marginal productivity:

$$Y_t(i) = A_t(i) H_t(i)^{\frac{1}{1+\eta}} \tag{4.27}$$

where $\eta > 0$. Each firm's productivity, $A_t(i)$, is given by

$$\ln(A_t(i)) = \ln(A(i)^*) + \epsilon_t^A + v_t^A(i) \tag{4.28}$$

where $A(i)^*$ is firm i's intrinsic productivity, while ϵ_t^A and $v_t^A(i)$ are mean zero aggregate and idiosyncratic shocks (each specified below) to the firm's productivity, broadly defined. Firm i's nominal marginal cost is then

$$MC_t(i) = (1+\eta) \frac{W_t}{A_t(i)^{1+\eta}} Y_t(i)^\eta \tag{4.29}$$

so that η is the elasticity of marginal cost with respect to output. Shocks to $A_t(i)$ may therefore be considered a reduced-form means of capturing shocks to firms' marginal costs other than those that act through demand or the wage. When combined with market clearing (see below), we can replace $Y_t(i)$ with the household's quantity demanded (4.24) to give

$$MC_t(i) = (1+\eta) \frac{W_t}{A_t(i)^{1+\eta}} \left(\frac{s_t(i) \mathbb{P}_t Y_t}{P_t(i)} \right)^\eta \tag{4.30}$$

Price setting

Although it remains our opinion that nominal rigidities represent the basis of *some* aggregate persistence, we here suppose that all firms are free to costlessly adjust their prices at the start of every period in order to highlight the real rigidities embodied in the current model. Firms' optimal price-setting rules under TransLog preferences will be examined in detail in sections 4.4 and 4.5 below.

4.3.3 Market clearing

All markets clear each period, so that:

$$\begin{aligned} &Y_t(i) = C_t(i) \quad && \forall t, i \\ &H_t = \int H_t(i)\, di \quad && \forall t \end{aligned} \tag{4.31}$$

This implies that aggregate output is given by:

$$C_t = Y_t = Z_t H_t^{\frac{1}{1+\eta}} \tag{4.32}$$

where aggregate TFP, Z_t, combines individual firm productivities, household preferences and a distortion from relative prices

$$Z_t \equiv \left(\int \left(\frac{A_t(i)}{s_t(i)} \frac{P_t(i)}{\mathbb{P}_t} \right)^{-(1+\eta)} di \right)^{-\frac{1}{1+\eta}} \tag{4.33}$$

4.3.4 The central bank

To close the model, we assume that the central bank sets nominal interest rates according to the Taylor-like policy function

$$\widehat{i}_t = \kappa_y E_t^{CB} [\widehat{y}_t] + \kappa_\pi E_t^{CB} [\widehat{\pi}_{t+1}] + \epsilon_t^M \tag{4.34}$$

where variables with a hat are deviations from steady-state, $E_t^{CB}[\cdot]$ is the mathematical expectation conditional on the central bank's information set in period t (defined below) and ϵ_t^M is a persistent, mean zero shock to monetary policy (specified below). Note that the component against inflation is against expected future inflation rather than current inflation, to provide a more accurate characterisation of modern central banking practice.

4.3.5 Information and timing

The representative household and the central bank are assumed to possess full information at all times, so that

$$E_t^{HH}[\cdot] = E_t^{CB}[\cdot] = E[\cdot|\Omega_t] \tag{4.35}$$

where Ω_t is the set of all information that exists in period t. In contrast, firms have incomplete information, so that

$$E_t(i)[\cdot] = E[\cdot|\mathcal{I}_t(i)] \tag{4.36}$$

where each period, firm i observes its own output from the previous period; its own productivity for the current period; and common, but imperfect signals for the previous period's aggregate price level and aggregate output:

$$\mathcal{I}_t(i) = \{\mathcal{I}_{t-1}(i), Y_{t-1}(i), A_t(i), \mathbb{P}_{t-1}e^{e_{p,t}}, Y_{t-1}e^{e_{y,t}}\} \tag{4.37}$$

Timing

Each period obeys the following timing

1. Innovations are drawn.
2. Firms observe $\mathcal{I}_t(i)$ and set their prices simultaneously.
3. The representative household and the central bank observe the full state of the economy and determine the interest rate, the real wage and the quantities of goods demanded for the given prices.
4. Firms produce the goods and the representative household consumes them.

Note, in particular, that firms do not observe any changes in households' relative preferences or the composition of their productivity shock before setting their prices.

4.3.6 Stochastic processes

The model contains uncertainty in the form of aggregate and idiosyncratic shocks on both the supply and demand sides of the economy, plus measurement error in aggregate statistics. We suppose that underlying aggregate shocks are persistent and follow AR(1) processes with Gaussian innovations.

$$\epsilon_t^A = \rho_\epsilon^A \epsilon_{t-1}^A + u_t^A \quad \text{where} \quad u_t^A \sim N\left(0, \sigma_{\epsilon^A}^2\right) \tag{4.38a}$$

$$\epsilon_t^M = \rho_\epsilon^M \epsilon_{t-1}^M + u_t^M \qquad u_t^M \sim N\left(0, \sigma_{\epsilon^M}^2\right) \tag{4.38b}$$

Measurement errors are transitory and Gaussian.

$$e_{p,t} \sim N\left(0, \sigma_{v^e}^2\right) \tag{4.38c}$$

$$e_{y,t} \sim N\left(0, \sigma_{v^e}^2\right) \tag{4.38d}$$

Idiosyncratic shocks are assumed to be transitory. Those to productivity are assumed to be Gaussian

$$v_t^A(i) \sim N\left(0, \sigma_{v^A}^2\right) \tag{4.38e}$$

while those to demand are left unspecified, except to note that they must satisfy

$$\begin{aligned} E\left[v_t^\alpha(i)\right] &= 0 \ \forall i, t \\ Var\left(v_t^\alpha(i)\right) &= \sigma_{v^\alpha}^2 \ \forall i, t \\ \alpha_t(i) &\in (0,1) \ \forall i, t \\ \sum\nolimits_i \alpha_t(i) &= 1 \ \forall t \end{aligned} \tag{4.38f}$$

All innovations are assumed to be fully independent from each other, both contemporaneously and across time.

Aggregated idiosyncratic shocks

As will be shown below, the following two linear aggregations of idiosyncratic shocks also enter into the model:

$$\widetilde{v}_t^A \equiv \sum\nolimits_i \alpha(i)^* v_t^A(i) \tag{4.39a}$$

$$\widetilde{v}_t^\alpha \equiv \sum\nolimits_i \alpha(i)^* v_t^\alpha(i) \tag{4.39b}$$

Since $\alpha(i)^* \in (0,1) \ \forall i$ and $\sum_i \alpha(i)^* = 1$, these are weighted averages of firms' idiosyncratic shocks. These statistics will not, in general, converge to zero as $J \to \infty$

because of the unequal weights applied across the firms' shocks. As discussed in section 1.3 of chapter 1, a power law distribution in the weights is sufficient to ensure that the Law of Large Numbers will not hold and we do indeed observe a power law distribution in firm sizes in the data.

4.3.7 Steady-state

With firms already possessing full flexibility in their price setting, we define a steady-state equilibrium to be that which holds when (a) there are no shocks to the system; and (b) this fact is common knowledge (i.e. a special case of firms also having full information).

Although not used in partial equilibrium – i.e. when considering only firms' prices, taking aggregate demand and wages as given – the following assumptions of *near-uniformity in steady-state preferences* will be required when considering general equilibrium:

$$\gamma_{ii} = -\frac{\gamma}{J} \, \forall i \; ; \; \gamma_{ij} = \frac{\gamma}{J\left(J-1\right)} \, \forall j \neq i \tag{4.40a}$$

$$\overline{p^*} = \frac{1}{J-1} \sum\nolimits_{j \neq i} p\left(j\right)^* = 0 \, \forall i \tag{4.40b}$$

$$s\left(i\right)^* = \frac{1}{J} \, \forall i \tag{4.40c}$$

It is helpful to explicitly enumerate what is and what is not uniform across goods under these assumptions. Things that *are* uniform across goods:

- Firms' consideration of price-competition: Γ
- Steady-state expenditure shares: $s\left(i\right)^* = \frac{1}{J}$
- Steady-state mark-ups: $\mu_i^* = 1 - \frac{s(i)^*}{\gamma_{ii}} = 1 + \frac{1}{\gamma}$

Things that are *not* uniform across goods:

- Steady-state base market shares: $\alpha\left(i\right)^*$
- Steady-state prices: $p\left(i\right)^*$
- Steady-state marginal costs: $MC\left(i\right)^*$

In other words, we capture a world in which, *even in steady-state*, there exist low-price, high-volume businesses operating next to high-price, low-volume businesses.

Applying (4.40a) - (4.40c) to the definition of expenditure share (4.14) gives a simple expression for each firm's steady-state price:

$$p\left(i\right)^{*}=\frac{1}{\gamma}\left(\zeta_{i}-1\right) \tag{4.41}$$

The RHS expresses, as a percentage, how far firm i's price is above the simple-average price when in steady-state. Plugging this into equation (4.5), we see that the log of the aggregate price level in steady-state is zero (this emerges from the normalisation that the simple-average price be zero):

$$\mathbb{p}^{*}\equiv\ln\left(\mathbb{P}^{*}\right)=0 \tag{4.42}$$

With a constant steady-state mark-up, pinning down a firm's price must also pin down its marginal cost. Our assumption of near-uniformity therefore gives a direct mapping between a firm's steady-state base market share, ζ_i, and its steady-state productivity, $A\left(i\right)^{*}$:

$$\ln\left(A\left(i\right)^{*}\right)=\frac{1}{1+\eta}\ln\left[\left(1+\frac{1}{\gamma}\right)\left(1+\eta\right)W^{*}\left(\frac{1}{J}Y^{*}\right)^{\eta}\right]-\frac{1}{\gamma}\left(\zeta_{i}-1\right) \tag{4.43}$$

At first glance, this might appear to suggest that firms with higher base market share are less productive. However, it should be remembered that $A_t\left(i\right)$ here should be construed as capturing all factors of production other than labour. A low value of $A\left(i\right)^{*}$ in this context is a shorthand means of saying that firm i has a lot of non-labour costs.

4.4 Price-setting under full information

Because firms are free to update their prices in every period, we can limit our attention to the one-period profit function:

$$\Pi_t(i) = s_t(i) N_t - C\left(\frac{s_t(i) N_t}{P_t(i)}\right) = s_t(i) N_t \left(1 - \frac{MC_t(i)}{P_t(i)}\right) \tag{4.44}$$

When firms have access to full information, there is no need to consider their expectations and the profit function can be maximised directly to give the usual expression of price as a mark-up over marginal costs:

$$\frac{P_t(i)}{MC_t(i)} = \mu_{i,t} = \frac{\varepsilon_{ii,t}}{\varepsilon_{ii,t} - 1} \tag{4.45}$$

However, this only pins down $P_t(i)$ *implicitly* in this context as the mark-up is endogenous under TransLog preferences. As the firm's price rises ($P_t(i) \uparrow$), its market share falls ($s_t(i) \downarrow$), driving its own-price elasticity higher ($\varepsilon_{ii,t} \uparrow$) and, hence, their optimal mark-up lower ($\mu_{i,t} \downarrow$).

Instead, appendix 4.A.2 provides a derivation of the following explicit solution for price-setting under TransLog preferences.

Proposition 3. *When monopolistically competitive firms face a system of demand characterised by TransLog preferences, conditional on each firm's share of household expenditure remaining within $s_t(i) \in (0,1)$, the optimal one-period, full-information price is given by*

$$\frac{P_t(i)}{MC_t(i)} = \nu_{i,t} = \mathcal{W}\left(\frac{e^{\phi_t(i)}}{MC_t(i)}\right) = \omega\left(\phi_t(i) - \ln\left(MC_t(i)\right)\right) \tag{4.46}$$

where $\phi_t(i)$ is defined as

$$\phi_t(i) \equiv 1 - \frac{1}{\gamma_{ii}}\left(\alpha_t(i) + \sum\nolimits_{j \neq i} \gamma_{ij} \ln\left(P_t(j)\right)\right) \tag{4.47}$$

Furthermore, when all goods are gross substitutes, there exists a unique positive, globally stable Nash equilibrium in prices, $\boldsymbol{P}_t^* = \boldsymbol{P}^*(\boldsymbol{\alpha}_t, \boldsymbol{MC}_t; \Gamma)$, that may be found by iterating through (4.46) - (4.47) from any non-zero initial price vector.

$\mathcal{W}(\cdot)$ is the Lambert $\mathcal{W}$ function, defined as the inverse of $f(\mathcal{W}) = \mathcal{W}e^{\mathcal{W}}$; and $\omega(\cdot)$ is the Wright ω function, defined as $\omega(x) \equiv \mathcal{W}(e^x)$. Section 4.4.1 below provides brief descriptions of these two functions.

The adjusted mark-up associated with the optimal price is no longer a simple expression of the firm's elasticity of demand, but is instead a function of the firm's marginal cost (i.e. the mark-up over marginal costs is itself a function of those marginal costs) and other firms' prices. In particular, since $\mathcal{W}(\cdot)$ is strictly increasing, (4.46) makes clear that in the absence of concurrent price increases by its competitors, a firm's optimal mark-up declines as its marginal cost increases. Likewise, $\phi(i)$ and, hence, the adjusted mark-up, is increasing in firm i's base market share and a weighted sum of its competitors' prices.

Note that $\sum_{j\neq i}(-\gamma_{ij}/\gamma_{ii}) = 1$, so that $\phi_t(i)$ contains a firm-specific weighted average of other firms' prices. $\phi_t(i)$ is related to the firm's endogenous mark-up, $\mu_{i,t}$, in the following manner

$$\phi_t(i) = \underbrace{1 - \frac{s_t(i)}{\gamma_{ii}}}_{\mu_{i,t}} + \ln(P_t(i)) \tag{4.48}$$

Consequently, we can use (4.48) to identify the prices at no-exit boundaries:

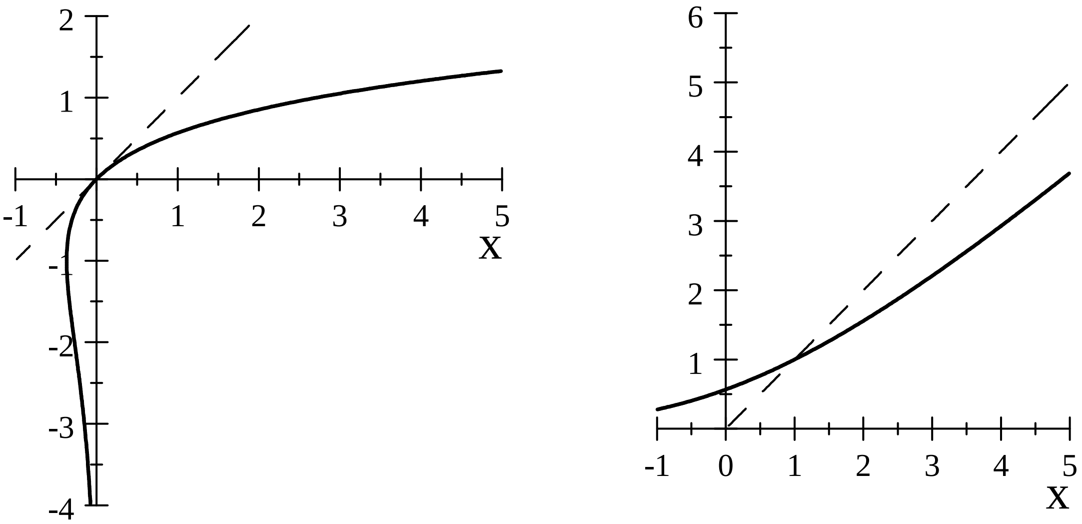
$$P_t(i)\big|_{s_t(i)=1} = e^{\phi_t(i)-1+\frac{1}{\gamma_{ii}}} \tag{4.49a}$$

$$P_t(i)\big|_{s_t(i)=0} = e^{\phi_t(i)-1} \tag{4.49b}$$

If a firm's price is too high, its share of household expenditure will fall to zero (i.e. it will exit the market), while if its price is too low, it will capture the entire market (i.e. force other firms to exit). That TransLog preferences include an endogenous exit rule is independently interesting, but we rule these possibilities out by assumption. In essence, this amounts to assuming that shocks are not too large.

4.4.1 The Lambert $\mathcal{W}$ and Wright ω functions

Illustrative plots of $\mathcal{W}(x)$ and $\omega(x)$ are provided in figure 4.1.



(a) Lambert $\mathcal{W}$

(b) Wright ω

Figure 4.1: The Lambert $\mathcal{W}$ and Wright ω functions in the real domain. Both plots include the 45° line for reference.

The Lambert $\mathcal{W}$ function

The Lambert $\mathcal{W}$ function, sometimes called the Omega function or the product logarithm, is defined as the inverse function of $f(\mathcal{W}) = \mathcal{W}e^{\mathcal{W}}$. For all $x \in \mathbb{R}_+$, $\mathcal{W}(x)$ is continuous, single valued, weakly positive, strictly increasing and concave. It has key values of $\mathcal{W}(0) = 0$ and $\mathcal{W}(e) = 1$; and its first derivative is given by $\frac{d\mathcal{W}(x)}{dx} = \frac{\mathcal{W}(x)}{x(1+\mathcal{W}(x))}$ for $x \notin \left\{0, -\frac{1}{e}\right\}$, with $\left.\frac{d\mathcal{W}(x)}{dx}\right|_{x=0} = 1$.

The Wright ω function

The Wright ω function, defined for $x \in \mathbb{R}$ as $\omega(x) \equiv \mathcal{W}(e^x)$, was first introduced and its properties discussed at length by Corless and Jeffrey (2002). For all $x \in \mathbb{R}_+$, $\omega(x)$ is continuous, single valued, strictly positive, strictly increasing and convex. It has key values of $\omega(0) = \mathcal{W}(1) \approx 0.56714$ and $\omega(1) = 1$. For $x \geq 1$, it lies beneath the 45° line and its first derivative is given by $\frac{d\omega(x)}{dx} = \frac{\omega(x)}{1+\omega(x)}$, so that $\lim\limits_{x\to\infty} \omega'(x) = 1$.

4.4.2 The optimal price as best response

Equations (4.46) - (4.47) express the optimal money price for firm i as a function of three objects – its nominal marginal cost, its base market share and a weighted average of other firms' prices. Combining these equations, we can rewrite the optimal price as:

$$\frac{P_t(i)}{MC_t(i)} = \nu_{i,t} = \mathcal{W}\left(e^{1-\frac{\alpha_t(i)}{\gamma_{ii}}}\prod_{j\neq i}\left(\frac{P_t(j)}{MC_t(i)}\right)^{\frac{-\gamma_{ij}}{\gamma_{ii}}}\right) \tag{4.50}$$

This formulation of price as a best response function makes clear that firm i's optimal mark-up is a function of its competitors' prices *relative to i's own marginal cost*.

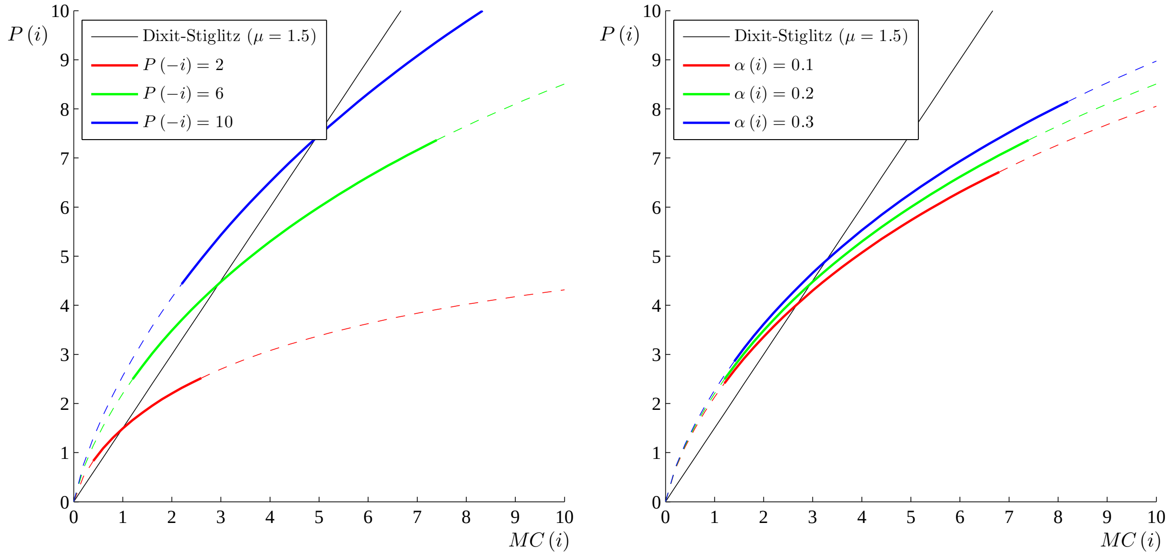
It is informative to consider the shape of this decision rule, as it helps understand firms' out-of-equilibrium pricing behaviour and so offers a precursor to section 4.5, where we consider price-setting under uncertainty. First, we have that the optimal price is strictly increasing in all three inputs:

$$\begin{aligned}
\frac{\partial P_t(i)}{\partial MC_t(i)} &= \frac{(\nu_{i,t})^2}{1+\nu_{i,t}} > 0 \\
\frac{\partial P_t(i)}{\partial P_t(j)} &= \left(\frac{\nu_{i,t}}{1+\nu_{i,t}}\right)\frac{MC(i)}{P(j)}\left(-\frac{\gamma_{ij}}{\gamma_{ii}}\right) > 0 \\
\frac{\partial P_t(i)}{\partial \alpha_t(i)} &= \left(\frac{\nu_{i,t}}{1+\nu_{i,t}}\right)MC(i)\left(-\frac{1}{\gamma_{ii}}\right) > 0
\end{aligned}$$

Note that if γ_{ii} is decreasing in the number of goods, as embodied in the uniformity-in-Γ restriction of (4.26b), then the sensitivity of price to base market share will be increasing in the same. When the number of goods is large, small fluctuations in consumers' relative preferences can have large effects on optimal prices. Second, we have that the optimal price is strictly concave in marginal cost and other prices, but strictly convex in base market share; and that the cross-derivatives are all strictly positive (full details of these derivatives may be found in appendix 4.A.2):

$$\begin{aligned}
\frac{\partial^2 P_t(i)}{\partial MC_t(i)^2} &< 0 & \qquad \frac{\partial^2 P_t(i)}{\partial MC_t(i)\,\partial P_t(j)} &> 0 \\
\frac{\partial^2 P_t(i)}{\partial P_t(j)^2} &< 0 & \qquad \frac{\partial^2 P_t(i)}{\partial MC_t(i)\,\partial \alpha_t(i)} &> 0 \\
\frac{\partial^2 P_t(i)}{\partial \alpha_t(i)^2} &> 0 & \qquad \frac{\partial^2 P_t(i)}{\partial P_t(j)\,\partial \alpha_t(i)} &> 0
\end{aligned}$$

These emphasise the importance of *co-ordination* in firms' price-setting decisions. A firm's ability to raise its price with its marginal cost is contingent on a concomitant rise in its competitors' prices, although this is partially mitigated if the firm commands a larger base market share. This feature of TransLog preferences is further illustrated in figures 4.2, 4.3 and 4.4.



(a) Cross-section by other firms' prices. $\alpha(i) = 0.2, \gamma_{ii} = -1$

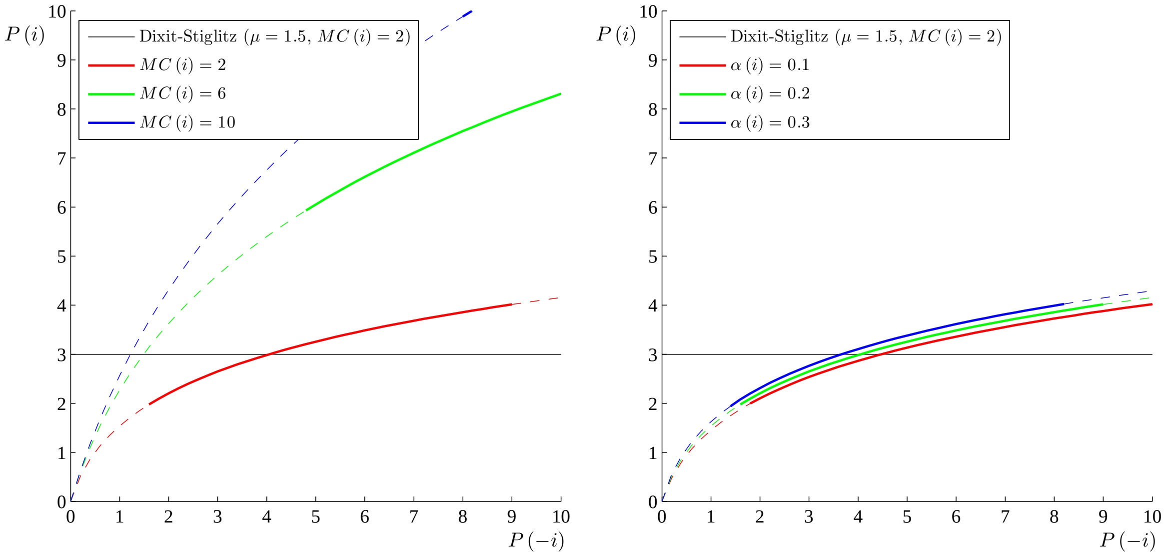
(b) Cross-section by base market share. $P(-i) = 6, \gamma_{ii} = -1$

Figure 4.2: Optimal price under full-information by nominal marginal cost.

Figure 4.2 plots firm i's full-information price as a function of its nominal marginal cost, with cross-sections shown by other firms' prices and base market share. Segments shown as solid lines are those in which the firm obtains an expenditure share bounded by 0 and 1, while segments shown as dashed lines fall outside these bounds (too low a price and the firm captures all household expenditure, too high a price and its share falls to zero). By way of comparison, the optimal price under Dixit-Stiglitz preferences is also shown, assuming an elasticity of $\varepsilon = 3$.

The concavity of the price-setting rule implies that the optimal price is sometimes below and sometimes *above* that suggested by the Dixit-Stigliz framework, depending on the marginal cost. With a flatter slope than the constant mark-up setting, this figure provides a simple illustration of the real rigidity embodied in

TransLog (and AIDS) preferences. Events that lower a firm's marginal cost, such as productivity improvements, will not cause that firm to lower their price so far as it would under a constant mark-up scheme. Likewise, the best response to inflation in a firm's input prices (such as wages) is to raise its output price by less than under a constant mark-up scheme when taking other firms' prices as given because of concerns regarding strategic complementarity. Indeed, for firms that receive a very low fraction of household expenditure, increases in marginal cost in the absence of increases in competitors' prices will cause a less than one-for-one increase in prices $\left(\frac{\partial P_t(i)}{\partial MC_t(i)} < 1\right)$.



(a) Cross-section by nominal marginal cost. $\alpha(i) = 0.2, \gamma_{ii} = -1$

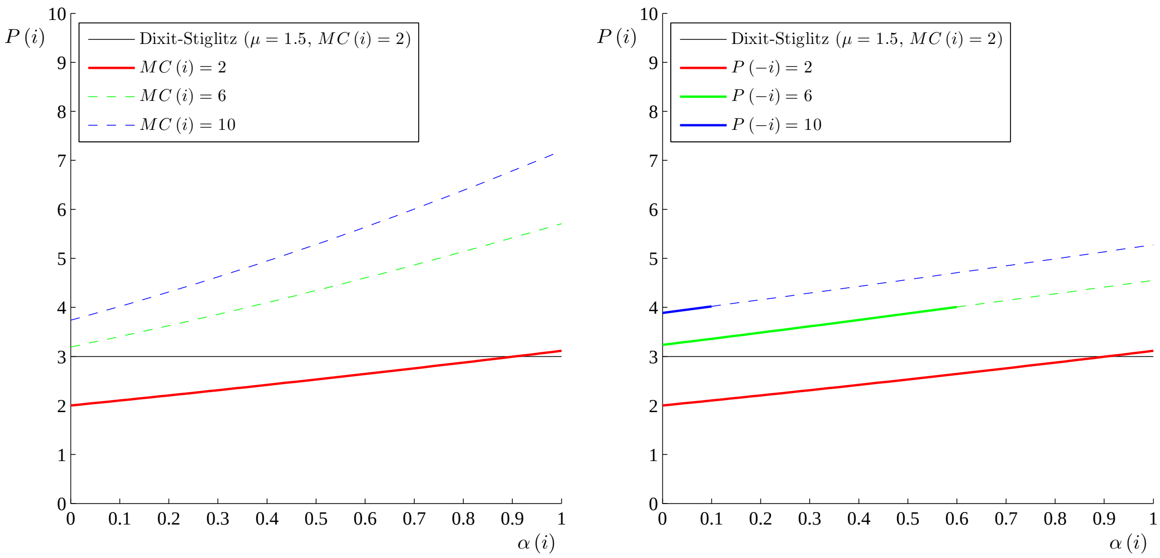
(b) Cross-section by base market share. $MC(i) = 2, \gamma_{ii} = -1$

Figure 4.3: Optimal price under full-information by a weighted average of other firms' prices.

Next, figure 4.3 shows that optimal prices are concave in other firms' prices and quite strongly so when marginal costs are low. Note, for example, that when $MC(i) = 2$, a price of 3 (i.e. a mark-up of 1.5) is only achieved when other firms' prices are roughly 4.

It is worth emphasising again that this strategic complementarity is firm-specific, with firm i being interested in the weighted-average $\ln(P(-i)) = \sum_{j \neq i} (-\gamma_{ij}/\gamma_{ii}) \ln(P(j))$

of its competitors' prices.[12] This implies a reduced opportunity for policy makers to assist coordination by publishing aggregate price statistics. With each firm interested in a different weighted sum, the aggregate price level will, at best, act as an imperfect signal to each firm's pricing problem.



(a) Cross-section by nominal marginal cost. $P(-i) = 2, \gamma_{ii} = -1$

(b) Cross-section by other firms' prices. $MC(i) = 2, \gamma_{ii} = -1$

Figure 4.4: Optimal price under full-information by base market share.

Finally, figure 4.4 shows that a firm's optimal price increases convexly in its base market share, but only quite weakly so. This convexity is strongest when marginal costs are high relative to other firms' prices, but doing so very quickly approaches the $s_t(i) = 0$ boundary. In other words, an increasing base market share allows a firm to raise its price, but it is generally not enough to offset falls in (or failures to increase) other firms' prices or its own marginal cost.

[12] Recall that the Γ matrix is symmetric and each row sums to zero, but there is no requirement that off-diagonal elements be equal. Indeed, such a case would be highly unusual.

4.4.3 Equilibrium prices under full information

Although each firm's best response to a common shock to marginal costs is to temper any price increase by lowering its mark-up, it will typically be the case that the *equilibrium* mark-up under full information is independent of the aggregate components of marginal cost. To see this, rewrite equation (4.50) as

$$\nu_{i,t} = \mathcal{W}\left(e^{1-\frac{\alpha_t(i)}{\gamma_{ii}}}\prod_{j\neq i}\left(\frac{MC_t\,(j)}{MC_t\,(i)}\nu_{j,t}\right)^{\frac{-\gamma_{ij}}{\gamma_{ii}}}\right)$$

Recall that a firm's nominal marginal cost (4.30) is given by

$$MC_t\,(i) = (1+\eta)\,\frac{W_t}{A_t\,(i)^{1+\eta}}\left(\frac{s_t\,(i)\,\mathbb{P}_t Y_t}{P_t\,(i)}\right)^{\eta}$$

Because it is the *ratio* of firms' marginal costs that matter for the equilibrium mark-up, all of the common components – wages, aggregate productivity and aggregate demand – will necessarily cancel out so that

$$\frac{MC_t\,(j)}{MC_t\,(i)} = \left(\frac{A_t\,(i)}{A_t\,(j)}\right)^{1+\eta}\left(\frac{s_t\,(j)\,P_t\,(i)}{s_t\,(i)\,P_t\,(j)}\right)^{\eta}$$

4.5 Price-setting under uncertainty

Under uncertainty, firm i's problem is to choose $P_t(i)$ to maximise its expected one-period profit, taking aggregate expenditure and other firms' prices as given

$$\max_{P_t(i)} E_t(i)\left[s_t(i)\,N_t\left(1-\frac{MC_t(i)}{P_t(i)}\right)\right] \tag{4.51}$$

where $E(i)[\cdot] \equiv E[\cdot|\mathcal{I}(i)]$ is the mathematical expectation conditional on information available to firm i at the start of the period, defined in section 4.3.5. Maintaining the notation that lower-case letters are the natural log of their upper-case counterparts, the corresponding first order condition (a derivation of this and the linear approximation below is provided in appendix 4.A.3) is

$$E_t(i)\left[\gamma_{ii}e^{n_t}\right] = E_t(i)\left[\gamma_{ii}e^{n_t+mc_t(i)-p_t(i)}\left(\phi_t(i)-p_t(i)\right)\right]$$

Denoting variables with an asterisk as being that variable in steady-state and variables with a hat above them to be that variable's deviation from steady-state – e.g., $\widehat{p}_t(i) \equiv p_t(i) - p(i)^*$ – we can construct a first-order Taylor series approximation of firm i's pricing rule around the no-shock, full-information equilibrium. Unlike the rule for the canonical Dixit-Stiglitz model – $\widehat{p}_t(i) = E_t(i)\left[\widehat{mc}_t(i)\right]$ – the TransLog system's pricing rule is:

$$\begin{aligned}\widehat{p}_t(i) = {} & \left(1-\frac{1}{1+\mu_i^*}\right)E_t(i)\left[\widehat{mc}_t(i)\right] \\ & + \left(\frac{1}{1+\mu_i^*}\right)E_t(i)\underbrace{\left[\left(\frac{-1}{\gamma_{ii}}\right)\widehat{\alpha}_t(i) + \sum_{j\neq i}\left(\frac{-\gamma_{ij}}{\gamma_{ii}}\right)\widehat{p}_t(j)\right]}_{=\widehat{\phi}_t(i)}\end{aligned} \tag{4.52}$$

where $\mu_i^* = 1 - \frac{s(i)^*}{\gamma_{ii}}$ is the mark-up employed by firm i in steady-state. A number of interesting results emerge from this pricing rule.

First, and most obviously, a firm's price is increasing in both its marginal cost and its mark-up, with the mark-up contribution being a positive combination of its base market share and its competitors' prices. That each firm's price increases in its base market share is not surprising. An increase in $\widehat{\alpha}_t(i)$ represents a shift in households' relative preferences towards good i, which grants the producer greater pricing power. Note that this variable is scaled by γ_{ii}, the parameter governing the price sensitivity of i's market share. If γ_{ii} is small, an increase in consumer preference for good i will have a large effect on its price.

Next, we see that for the purposes of setting its mark-up, firm i is interested in a *firm-specific* weighted average of competitors' prices and not the aggregate price level.[13] This inclusion of competitors' prices in the pricing equation is in addition to those brought in via marginal costs as would be typical in the constant-mark-up setting.

Finally, since each firm's steady-state mark-up increases with its market share, we have that larger firms will, *ceteris paribus*, place relatively more weight on movements in their marginal costs than on movements in their competitors' prices or consumer preferences. Similarly, smaller firms will pay more attention to price competition and consumer preferences, relative to their larger counterparts.

Of course, if a firm's marginal cost is increasing in the quantity of goods it produces, this acts as a further source of real rigidity no matter what the model of demand. Linearising and substituting the expression for firms' nominal marginal costs (4.30) into (4.52) and exploiting the transitory nature of idiosyncratic shocks to demand, we obtain:

$$\begin{aligned} \widehat{p}_t\left(i\right)^{TL} = E_t\left(i\right)&\left[\theta_i\,\widehat{\mathbb{p}}_t + \left(1-\theta_i\right)\sum_{j\neq i}\left(\frac{-\gamma_{ij}}{\gamma_{ii}}\right)\widehat{p}_t\left(j\right)\right] \\ &+ \left(\frac{\theta_i}{1+\eta}\right)E_t\left(i\right)\left[\widehat{\varpi}_t - \left(1+\eta\right)\widehat{a}_t\left(i\right) + \eta\widehat{y}_t\right] \end{aligned} \tag{4.53}$$

where $\widehat{\varpi}_t$ is the real wage and η is the elasticity of marginal cost with respect to output, while $\theta_i \equiv \left(\frac{\mu_i^*(1+\eta)}{1+\mu_i^*\left(1+\eta\varepsilon_{ii}^*\right)}\right) \in (0,1)$. Each firm will estimate three broad objects: the aggregate price level; a firm-specific weighted average of other firms' prices; and the real components of its marginal cost.

The nominal considerations include a component from the firm's marginal cost and a component from its mark-up, thereby clearly delineating firms' price concerns

[13] Recall that $\gamma_{ii} < 0$; $\gamma_{ij} \geq 0\ \forall i, j$; and $\sum_{j\neq i}\left(-\gamma_{ij}/\gamma_{ii}\right) = 1$.

with respect to its cost of doing business and with respect to the strategic complementarity of price competition. Since θ_i is increasing in $\alpha(i)^*$, we have that larger firms will focus more on the real determinants of marginal cost and the aggregate price level, while smaller firms will place more weight on price competition.

This pricing rule embodies two distinct sources of real rigidity above the standard Dixit-Stiglitz framework. First, it accounts for curvature in demand and the corresponding variation in firms' mark-ups. Second, it requires that firms estimate multiple aggregated price statistics (because of asymmetries in consumer preferences) in an environment where only one – the aggregate price level – is published.

It is possible to separate these two effects by examining the pricing rule with complete uniformity in preferences as a part-way point between Dixit-Stiglitz pricing and the more realistic TransLog environment. Imposing complete uniformity gives us that $\widehat{\mathbb{p}}_t = \sum_{j \neq i} \left(\frac{-\gamma_{ij}}{\gamma_{ii}}\right) \widehat{p}_t(j) = \overline{\widehat{p}_t}\ \forall i$, so that (4.53) may be written as

$$\widehat{p}_t(i)^{UTL} = E_t(i)\left[\overline{\widehat{p}_t}\right] + \left(\frac{\mu^*}{1+\mu^*(1+\eta\varepsilon^*)}\right) E_t(i)\left[\widehat{\varpi}_t - (1+\eta)\,\widehat{a}_t(i) + \eta\widehat{y}_t\right] \tag{4.54}$$

where the superscript "UTL" is used to denote "Uniform TransLog." By contrast, the linearised pricing rule under Dixit-Stiglitz preferences is

$$\widehat{p}_t(i)^{DS} = E_t(i)\left[\overline{\widehat{p}_t}\right] + \left(\frac{1}{1+\eta\varepsilon}\right) E_t(i)\left[\widehat{\varpi}_t - (1+\eta)\,\widehat{a}_t(i) + \eta\widehat{y}_t\right] \tag{4.55}$$

Firms are interested in the same estimating the same objects under both environments, but attribute less weight to the real components of marginal cost under TransLog preferences ($\frac{\mu}{1+\mu(1+\eta\varepsilon)} < \frac{1}{1+\eta\varepsilon}$), reflecting the fact that variable mark-ups reduce any response to changes in marginal cost.

4.5.1 Applying near-uniformity in preferences

The pricing rule of equation (4.53) is, in general, intractable when seeking to include it in a simulated general equilibrium model. This is because when aggregating (substituting 4.53 back into itself), it stipulates that firms consider $J+1$ different compound expectations: one for the aggregate price level plus a different one for every firm. As shown in chapter 1, the state vector will therefore explode in J (in addition to in the number of higher-orders).

To progress further, is is therefore necessary now to apply our assumption of near uniformity in preferences (4.40a) - (4.40c). These simplify the pricing rule (4.53) to

$$\begin{aligned}\widehat{p}_t\left(i\right)^{NUTL} = {} & E_t\left(i\right)\left[\theta\,\widehat{\mathbb{p}}_t + \left(1-\theta\right)\overline{\widehat{p}_t}\right] \\ & + \left(\frac{\mu^*}{1+\mu^*\left(1+\eta\varepsilon^*\right)}\right)E_t\left(i\right)\left[\widehat{\varpi}_t - \left(1+\eta\right)\widehat{a}_t\left(i\right) + \eta\widehat{y}_t\right]\end{aligned} \tag{4.56}$$

where the superscript "NUTL" is used to denote "Near-Uniform TransLog"; $\theta = \left(\frac{\mu^*(1+\eta)}{1+\mu^*(1+\eta\varepsilon^*)}\right)$; and $\frac{1}{J-1}\sum_{j\neq i}\widehat{p}_t\left(j\right) = \overline{\widehat{p}_t}\ \forall i$. In particular, we now have that every firm is interested in exactly the same two linear combinations of individual firms' prices. This unfortunately removes the result that large and small firms place different weights on supply and demand considerations (equality in steady-state expenditure shares means that the θ_is must be equal). Even so, it remains the case that since firms are interested in two different aggregate price statistics but observe only one, their ability to find the optimal price will be impaired relative to the STL case.

We show in appendix 4.A.4 that under near-uniformity, the linearised aggregate price level (4.5) is given by

$$\widehat{\mathbb{p}}_t = \widetilde{\widehat{p}}_t + \frac{1}{\gamma}\widetilde{v}_t^{\alpha} \tag{4.57}$$

where

$$\widetilde{\widehat{p}}_t \equiv \sum_i \alpha\left(i\right)^* \widehat{p}_t\left(i\right) \quad \text{and} \quad \widetilde{v}_t^{\alpha} \equiv \sum_i \alpha\left(i\right)^* v_t^{\alpha}\left(i\right)$$

which is to say that the aggregate price level is a weighted average of all prices plus a transitory aggregated preference shock.

We also show in appendix 4.A.4 that linearising the household's intratemporal first-order condition (4.22), aggregate production (4.32) and aggregate TFP (4.33), and substituting them into (4.56) then gives

$$\widehat{p}_t(i)^{NUTL} = \theta^* E_t(i)[\widehat{\mathbb{p}}_t] + (1-\theta^*) E_t(i)\left[\widehat{\overline{p}_t}\right] + E_t(i)\left[\theta\left(\frac{\frac{1}{\sigma}+\eta}{1+\eta}+\frac{1}{\psi}\right)\widehat{y}_t - \theta\left(1+\frac{1}{\psi}\right)\epsilon_t^A - \theta v_t^A(i)\right] \tag{4.58}$$

where $\theta^* = \theta\left(1+\frac{1}{\psi}\right)$. Note that an expectation regarding the idiosyncratic productivity shock remains here while that of the demand shock does not because it is something about which the firm *can* form an expectation. By observing the combined (aggregate plus idiosyncratic) shock to its costs, a firm can and will form an opinion on what of that is common to all firms and the remainder must, by definition, be their idiosyncratic component. By contrast, no information available to the firm allows them to move away from its *a priori* expectation of its idiosyncratic demand shock; namely, that it be zero.

4.5.2 Higher-order expectations

Linearising the household Eular equation (4.23) gives

$$\widehat{y}_t = E_t[\widehat{y}_{t+1}] - \sigma\left(\widehat{i}_t - E_t[\widehat{\mathbb{p}}_{t+1} - \widehat{\mathbb{p}}_t]\right) \tag{4.59}$$

We shown in appendix 4.A.5 that substituting (4.59), the central bank's policy rule (4.34), and the linearised aggregate price level (4.57) into (4.58), and recognising that $E_t(i)[v_t^\alpha(j)] = 0\ \forall j$, gives

$$\begin{aligned}\widehat{p}_t(i)^* = {} & \lambda_1^* E_t(i)\left[\widehat{\overline{p}_t}\right] + \lambda_2^* E_t(i)\left[\widehat{\widetilde{p}_t}\right] \\ & + \lambda_3^* E_t(i)[\widehat{\mathbb{p}}_{t+1}] + \lambda_4^* E_t(i)[\widehat{y}_{t+1}] \\ & + \lambda_5^* E_t(i)\left[\epsilon_t^A\right] + \lambda_6^* E_t(i)\left[\epsilon_t^M\right] \\ & + \lambda_7^* \widehat{a}_t(i)\end{aligned} \tag{4.60}$$

for $* \in \{NUTL, UTL, DS\}$. The λ_*^* coefficients are described in detail in the appendix, although we note here that under some mild restrictions on the central bank's policy function, we have

$$\lambda_1 \in (0,1)\,;\ \lambda_2 \in (0,1)\,;\ \text{and}\ \lambda_1+\lambda_2 \in (0,1)$$

These restrictions are necessary to ensure that decreasing weight is placed against higher-order expectations and, hence, to ensure that the system is stable.

Equation (4.60) describes each firm's pricing rule as a linear combination of two aggregate price statistics: a *weighted*-average price and a *simple*-average price.[14] We show in appendix 4.A.5 that if we take the simple average of equation (4.58) and repeatedly substitute it in for $\overline{\widehat{p}_t}$, we obtain

$$\widehat{p}_t\left(i\right)=E_t\left(i\right)\left[\sum_{k=0}^{\infty}\lambda_1^k\overline{E}_t^{(k)}\left[\lambda_2\widetilde{\widehat{p}_t}+\lambda_3\widehat{\mathbb{p}}_{t+1}+\lambda_4\widehat{y}_{t+1}+\left(\lambda_5+\lambda_1\lambda_7\right)\epsilon_t^A+\lambda_6\epsilon_t^M\right]\right] \\ +\lambda_7\,\widehat{a}_t\left(i\right) \tag{4.61}$$

where $\overline{E}_t^{(k)}\left[\cdot\right]$ is the k-th order simple-average expectation and we use the standard notation that the 0-th order expectation of an object is the object itself.

For Uniform TransLog and Dixit-Stiglitz preferences, we note that $\lambda_2=0$ and that the aggregate price level is just the simple-average price, so that

$$\widehat{\mathbb{p}}_t^*=\lambda_7^*\,\epsilon_t^A \\ +\sum_{k=0}^{\infty}\left(\lambda_1^*\right)^k\overline{E}_t^{(k+1)}\left[\lambda_3^*\widehat{\mathbb{p}}_{t+1}^*+\lambda_4^*\widehat{y}_{t+1}+\left(\lambda_5^*+\lambda_1^*\lambda_7^*\right)\epsilon_t^A+\lambda_6^*\,\epsilon_t^M\right] \tag{4.62}$$

for $*\in\{UTL,DS\}$.

For Near-Uniform TransLog preferences, on the other hand, we now take the weighted average of (4.61), repeatedly substitute it back in for $\widetilde{\widehat{p}_t}$ and then combine the result with the expression for the aggregate price level (4.57) to arrive at

$$\widehat{\mathbb{p}}_t^*=\lambda_7^*\left(\epsilon_t^A+\widetilde{v}_t^A\right)+\frac{1}{\gamma}\,\widetilde{v}_t^\alpha \\ +\widetilde{E}_t\left[\boldsymbol{\delta}'\mathbb{E}_t^{(0:\infty)}\left[\begin{array}{l}\lambda_3^*\widehat{\mathbb{p}}_{t+1}^*+\lambda_4\widehat{y}_{t+1}+\left(\lambda_5^*+\lambda_1^*\lambda_7^*+\lambda_2^*\lambda_7^*\right)\epsilon_t^A \\ +\lambda_6^*\epsilon_t^M+\lambda_2^*\lambda_7^*\widetilde{v}_t^A\end{array}\right]\right] \tag{4.63}$$

where $*=$ "NUTL"; $\mathbb{E}_t^{(0:\infty)}\left[\cdot\right]$ is the hierarchy of all permutations of the two compound expectations; and $\boldsymbol{\delta}$ assigns geometrically decreasing weights to each, with λ_1^*

[14] For TransLog preferences, the weighted-average price is for the aggregate price level (through the nominal marginal cost) and the simple-average price for price competition through the mark-up. For Symmetric TransLog and Dixit-Stiglitz preferences, only the simple-average is required.

raised to the power of the number of orders of $\overline{E}_t[\cdot]$ and λ_2^* raised to the power of the number of orders of $\widetilde{E}_t[\cdot]$.

For example, a weight of $(\lambda_1^*)(\lambda_2^*)^2$ is applied to both $\overline{E}_t\left[\widetilde{E}_t\left[\widetilde{E}_t[\cdot]\right]\right]$ and $\widetilde{E}_t\left[\overline{E}_t\left[\widetilde{E}_t[\cdot]\right]\right]$ as there are two orders of weighted-average expectations and one order of simple-average expectations in each.

4.5.3 Firms' learning

We define $\boldsymbol{x}_t$ to be the vector of aggregate shocks:

$$\boldsymbol{x}_t \equiv \begin{bmatrix} \epsilon_t^A \\ \epsilon_t^M \end{bmatrix} \tag{4.64}$$

and X_t as the full hierarchy of expectations regarding $\boldsymbol{x}_t$ formed in period t:

$$X_t \equiv \mathbb{E}_t^{(0:\infty)}[\boldsymbol{x}_t] = \begin{bmatrix} \boldsymbol{x}_t \\ \overline{E}_t[X_t] \\ \widetilde{E}_t[X_t] \end{bmatrix} \tag{4.65}$$

In the linearised model, firms receive the following vector of signals each period (see the definition of firms' information sets (4.37)):

$$\boldsymbol{q}_t(i) = \begin{bmatrix} \widehat{y}_{t-1} + e_{y,t} \\ \widehat{\mathbb{p}}_{t-1} + e_{p,t} \\ \widehat{y}_{t-1}(i) \\ \widehat{a}_t(i) \end{bmatrix} \tag{4.66}$$

Note that there is no need to include the firm's previous-period price as it must necessarily have been a function of $\mathcal{I}_{t-1}(i)$, meaning that it contains no new information in period t. The firm's previous-period quantity is still relevant, though, as firms do not observe the demand they face each period until after setting their prices. The period $t-1$ individual quantity demanded therefore contains news to a firm when setting its price in period t.

We show in appendix 4.A.7 that this signal vector may be written as

$$\boldsymbol{q}_t(i) = C_1 X_t + C_2 X_{t-1} + R_1 \begin{bmatrix} v_t^A(i) \\ v_t^\alpha(i) \end{bmatrix} + R_2 \begin{bmatrix} v_{t-1}^A(i) \\ v_{t-1}^\alpha(i) \end{bmatrix} + R_3 \boldsymbol{e}_t \tag{4.67}$$

Because of the linearity of the underlying system, the best linear estimator – in the sense of minimising the mean squared error – will be a Kalman filter:[15][16]

$$E_t(i)[X_t] = E_{t-1}(i)[X_t] + K\{\boldsymbol{q}_t(i) - E_{t-1}(i)[\boldsymbol{q}_t(i)]\} \tag{4.68}$$

In appendix 4.A.6 we derive expressions for the time-invariant Kalman gain matrix (K) and the corresponding variance-covariance matrix (V), making use of the techniques developed by Nimark (2011b) to account for firms' signal vectors (4.67) including lagged components without the need to stack the current state on the previous.

Taking simple and weighted averages of (4.68), we then show that the following law of motion for the hierarchy of firms' expectations emerges:

$$X_t = FX_{t-1} + G_1\boldsymbol{u}_t + G_2\widetilde{\boldsymbol{v}}_t + G_3\boldsymbol{e}_t \tag{4.69}$$

where

$$\boldsymbol{u}_t = \begin{bmatrix} u_t^A \\ u_t^M \end{bmatrix} \quad \widetilde{\boldsymbol{v}}_t = \begin{bmatrix} \widetilde{v}_t^A \\ \widetilde{v}_t^\alpha \end{bmatrix} \quad \boldsymbol{e}_t = \begin{bmatrix} e_{p,t} \\ e_{y,t} \end{bmatrix} \tag{4.70}$$

Note that the Kalman filter is performing two roles here. It both represents the rule by which firms update their expectations and defines the law of motion for the state vector of interest.

4.5.4 Solving the model

We now have that in addition to (4.69) describing the law of motion for the hierarchy of firms' expectations, the economy is characterised by the following expression for real GDP:

$$\widehat{y}_t = \frac{1}{1+\sigma\kappa_y}E_t[\widehat{y}_{t+1}] - \frac{\sigma}{1+\sigma\kappa_y}\left((\kappa_\pi - 1)E_t[\widehat{\mathbb{p}}_{t+1} - \widehat{\mathbb{p}}_t] + \epsilon_t^M\right)$$

[15] If all shocks were drawn from Gaussian distributions, it would be the best such estimator, linear or otherwise.

[16] The derivation of the standard Kalman filter may be found in most texts on dynamic macroeconomics (e.g. Ljungqvist and Sargent, 2004) or timeseries analysis (e.g. Hamilton, 1994).

and one of (4.62) or (4.63) for the price level, depending on which demand system is being used. We show in appendix 4.A.7 that this system may be written as

$$\widehat{y}_t^* = \boldsymbol{\gamma}_y^{*\prime} X_t + \boldsymbol{\delta}_y^{*\prime} \widetilde{\boldsymbol{v}}_t \tag{4.71a}$$

$$\widehat{\mathbb{p}}_t^* = \boldsymbol{\gamma}_p^{*\prime} X_t + \boldsymbol{\delta}_p^{*\prime} \widetilde{\boldsymbol{v}}_t \tag{4.71b}$$

$$\widehat{\overline{p}}_t^* = \boldsymbol{\gamma}_{\overline{p}}^{*\prime} X_t \tag{4.71c}$$

$$\widehat{w}_t^* = \boldsymbol{\gamma}_w^{*\prime} X_t + \boldsymbol{\delta}_w^{*\prime} \widetilde{\boldsymbol{v}}_t \tag{4.71d}$$

where $* \in \{DS, UTL, NUTL\}$. Note that under Dixit-Stiglitz and Uniform Trans-Log preferences, (a) the law of large number ensures that $\widetilde{\boldsymbol{v}}_t = 0 \ \forall t$ almost surely, and (b) the aggregate price level and the simple-average price are the same. The terms in $\widetilde{\boldsymbol{v}}_t$ and the distinction between the two aggregated prices only matter when preferences are not fully uniform across goods in steady-state.

4.6 Simulations

Table 4.1 lists baseline parameters for the simulation presented below. As in chapter 3, the coefficient against expected next-period inflation is less than unity in order to ensure the existence of a non-explosive solution.

Parameter	Value	Description
σ	0.5	The elasticity of intertemporal substitution
ψ	1.5	The Frisch elasticity of labour supply
ε_{ii}^{*}	3.0	The steady-state own-price elasticity of demand
ζ^{*}	0.063	The degree of asymmetry in steady-state base market shares
η	0.5	The elasticity of marginal cost ($\equiv \frac{\alpha}{1-\alpha}$, $\alpha = 0.333$)
κ_y	0.5	The CB's coefficient against current real GDP
κ_π	0.5	The CB's coefficient against expected next-period inflation
ρ_A	0.6	AR(1) coefficient for aggregate TFP shocks
ρ_M	0.6	AR(1) coefficient for monetary shocks
$\frac{\sigma_v^2}{\sigma_u^2}$	5	Relative volatility of idiosyncratic shocks

Table 4.1: Baseline parameterisation

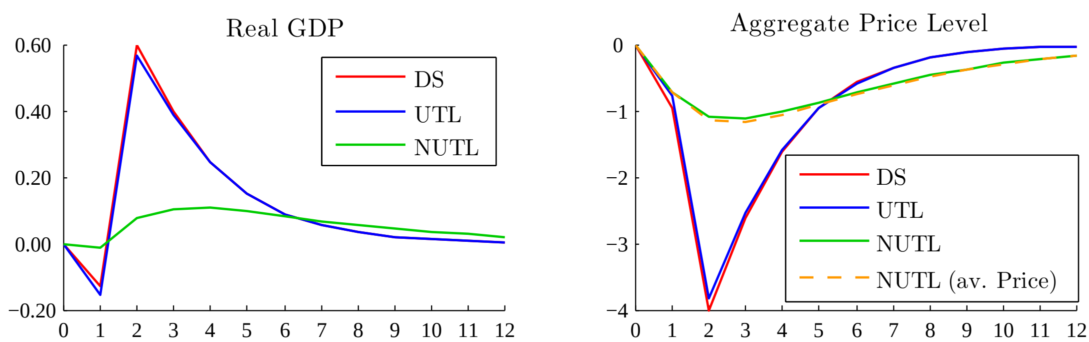
For the two TransLog demand systems, the steady-state own-price elasticity of demand ties down the deep parameter γ that controls cross-price elasticity ($\varepsilon_{ii}^{*} = 1 + \gamma$). For Near-Uniform TransLog preferences, $\zeta^{*} \equiv lim_{J\to\infty} \frac{1}{J} \sum_{i=1}^{J} (\zeta_i)^2$ characterises the degree of asymmetry in the distribution of steady-state base market shares. The value of 0.063 was chosen as this corresponds to a Zeta distribution with a shape parameter of $\gamma = 1.25$[17] and Axtell (2001) finds that the size of firms in the USA, when estimated over a 10 year period, are well characterised by such a distribution.

In everything that follows, period 0 denotes the period immediately prior to any shock occurring (the economy is invariably assumed to be in steady-state in period 0) and period 1 denotes the "on impact" period.

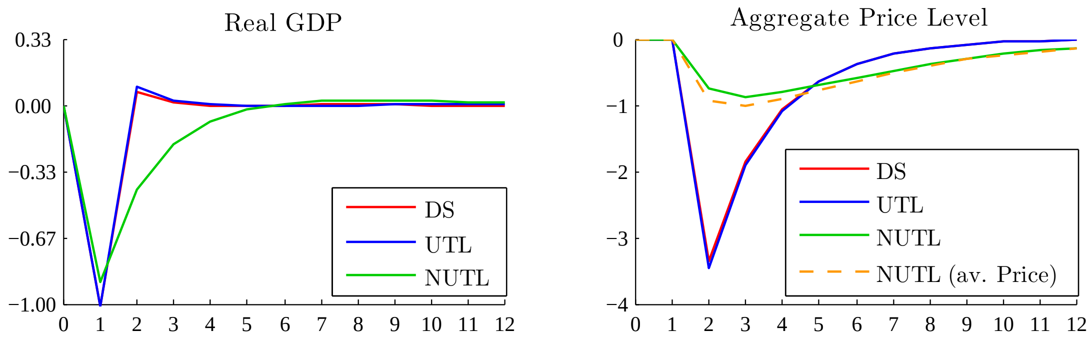
[17] See assumption 2 in chapter 2. The same logic applies here, with the p.d.f. $\phi_N(i)$ simply replaced with $\alpha_J(i)^{*}$.

4.6.1 Comparing demand systems

Figure 4.5 plots impulse responses for each of the three demand systems following one s.d. shocks to each of the two underlying state variables under the baseline parameterisation.



(a) Aggregate TFP shock



(b) Monetary Policy shock

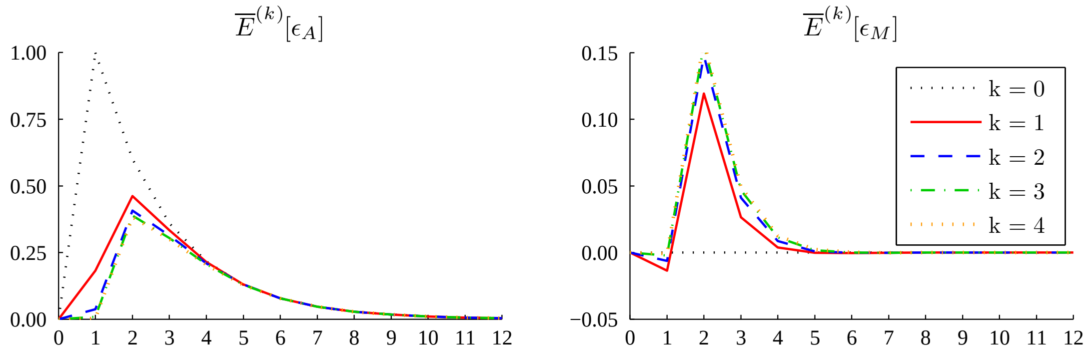
Figure 4.5: IRFs for the three systems of demand

The impulse responses for Dixit-Stiglitz and Uniform TransLog preferences are very similar for both aggregate shocks. For the shock to productivity, firms on impact attribute their private observations of productivity increases to idiosyncratic factors and consequently lower their prices only somewhat. In period 2, they observe the change in aggregate variables and their private demand that occurred in period 1 and immediately respond fully to the shock. Prices and real GDP then return to zero as the underlying shock itself dies away. For the shock to monetary policy, since firms observe no contemporary signal of household demand before setting their prices, there is no change in aggregate prices on impact and real GDP absorbs all of the effect. In period 2, upon observing aggregate signals and their own quantities from the impact period, firms' prices adjust fully and real GDP essentially returns to zero (the slight positive effect in period 2 arises from firms' average expectation that there is a slight increase in aggregate TFP as well - see figure 4.6a below). In other words, with firms having full flexibility in their price-setting, we essentially have the standard result of money being neutral, with monetary policy shocks having only a transitory real effect and prices capturing the aggregate effects thereafter.

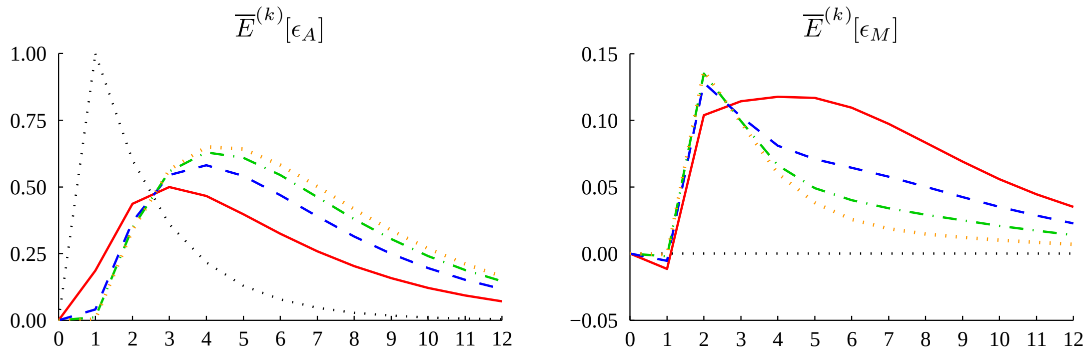
However, Near-Uniform TransLog preferences induce responses that stand in marked contrast. For the shock to productivity, the same 1 standard deviation innovation as under Dixit-Stiglitz and Uniform TransLog preferences produces responses that are more subdued, noticeably more hump-shaped and considerably more persistent than those of UTL or DS (so that relative to their peak response, NUTL preferences induce very much more persistent responses). The aggregate price level and the simple-average price move essentially one-for-one. For a monetary shock, despite the absence of any nominal rigidity and firms having access to the same information as under UTL or DS preferences, the neutrality of money result is lost: the subdued price response of NUTL grants the policy shock a considerably more persistent effect on real GDP.

To appreciate the increased persistence associated with NUTL preferences, we look at impulse responses for firms' hierarchies of simple-average expectations regarding the two underlying state variables following each of the two shocks. First, figure 4.6 plots those hierarchies for UTL and NUTL preferences following an aggregate TFP shock.

Under Uniform TransLog preferences, on impact, firms attribute the majority of their observed increases in productivity to idiosyncratic forces (because of the latter's higher variance). In period 2, the public signals regarding aggregate variables in period 1 increase the average belief considerably, but the measurement errors in those public signals mean that the average belief is still below the truth. Firms in period 2 also partially attribute the increase in real GDP to increased demand following a monetary policy shock, although this belief dies away very quickly over



(a) Uniform TransLog (UTL) preferences



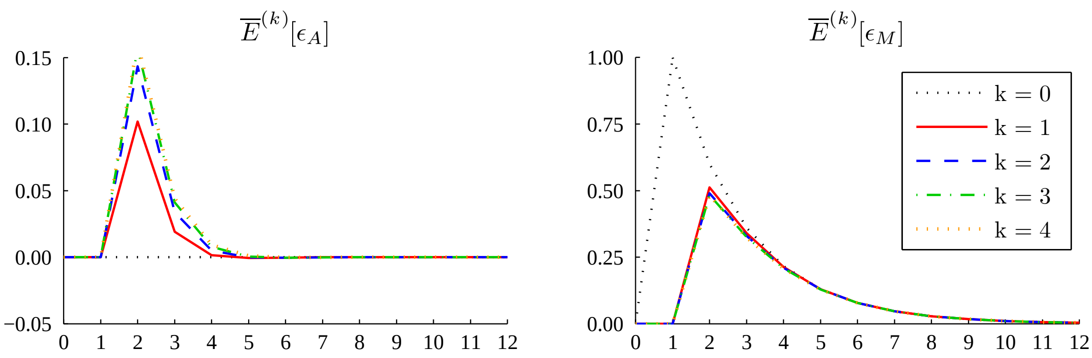
(b) Near-Uniform TransLog (NUTL) preferences

Figure 4.6: Hierarchies of simple-average expectations following a TFP shock

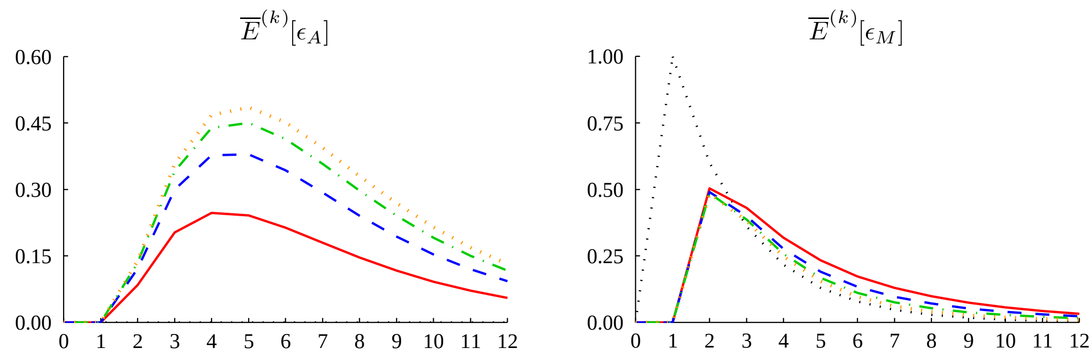
the next few periods. The impulse responses for aggregate beliefs under Dixit-Stiglitz preferences are broadly the same as those shown in figure 4.6a.

Under Near-Uniform TransLog preferences, the responses on impact are identical to those under UTL preferences, but thereafter differ considerably. Because firms must estimate both the aggregate price level and the simple-average price, but only receive public signals regarding the former, they incorrectly attribute some of the movements in their signals to shocks in both the *level* and the *distribution* of demand. This leads average expectations regarding aggregate TFP to be higher and considerably more persistent than the truth, and those regarding monetary shocks to also be more persistent (although still quite small).

Figure 4.7 next plots the equivalent graphs following a monetary policy shock.



(a) Uniform TransLog (UTL) preferences



(b) Near-Uniform TransLog (NUTL) preferences

Figure 4.7: Hierarchies of simple-average expectations following a monetary policy shock

For UTL preferences (DS preferences are largely the same), the absence of any signal of aggregate demand before setting their prices in a given period means that firms' beliefs do not move on impact. In period 2, the observed signal of real GDP from period 1 raises their expectations regarding both underlying variables, but in period 3 the information they receive in period 2 partially confirms that a monetary policy shock has occurred and so lowers their estimates of aggregate TFP. Beliefs regarding the former then follow the truth back to zero, while the latter have returned to zero after a handful of periods.

For NUTL preferences, there is likewise no on-impact response, but additional persistence thereafter. Average expectations regarding aggregate TFP rise higher and persist throughout, while those regarding the monetary shock fall to zero from *above* the truth. Once again, these reflect the more challenging signal extraction problem faced by firms under NUTL preferences.

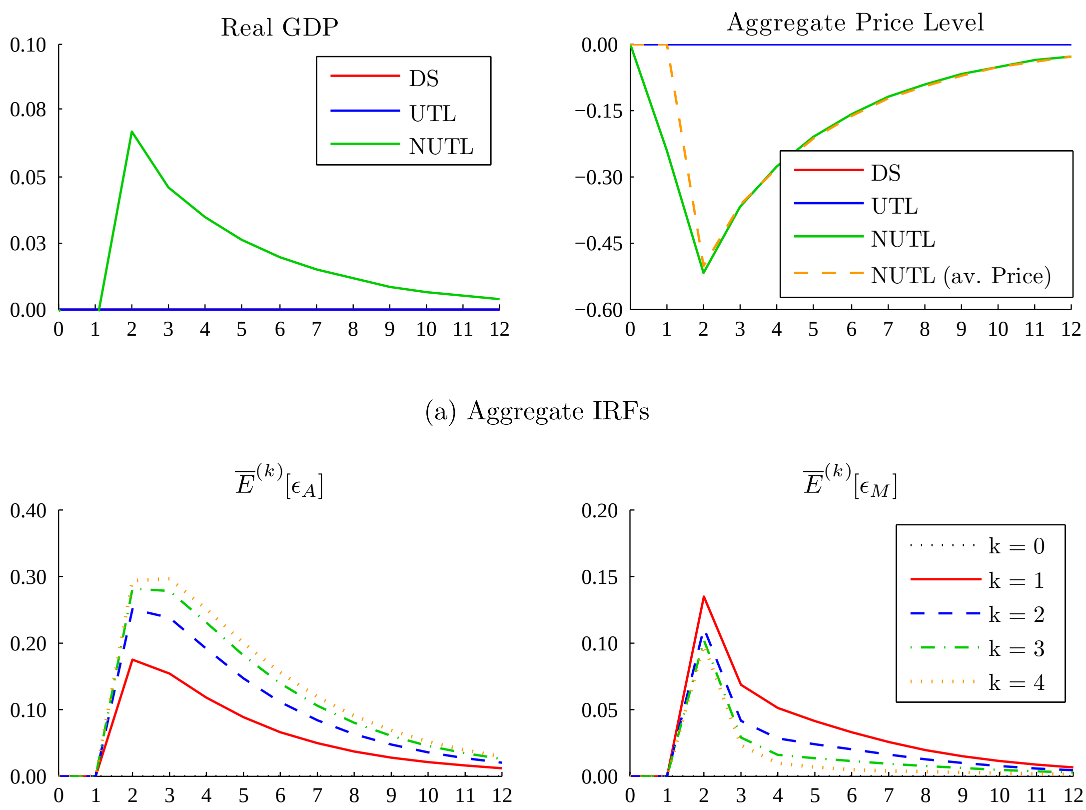
4.6.2 Aggregate volatility from idiosyncratic shocks under NUTL preferences

Under Near-Uniform TransLog preferences, the aggregate price level is comprised of the *weighted*-average of individual firms' prices and the weighted-average of firms' idiosyncratic shocks in that period, with the weights given by firms' steady-state base market shares. Since the weighted-average of idiosyncratic shocks need not be subject to the law of large numbers, NUTL preferences therefore give rise to aggregate volatility emerging from firms' idiosyncratic shocks.

To illustrate this, figures 4.8 and 4.9 plot aggregate impulse responses and the hierarchies of simple-average expectations following shocks to $\widetilde{v_t^A}$ and $\widetilde{v_t^\alpha}$ respectively. That is, following situations in which firms with relatively large steady-state marginal costs experience positive shocks while firms with relative small steady-state marginal costs experience negative shocks (recall that in steady-state all firms receive the same share of household expenditure; see section 4.3.7 for a description of steady-state).

A shock to more expensive firms' productivity (figure 4.8) induces aggregate responses that are suggestive of a true aggregate shock to productivity. The aggregate price (which is a weighted-average) falls on impact, but the simple-average price does not because the law of large numbers holds for it. Demand rises because of the fall in the aggregate price level. In period 2, all firms observe that real GDP rose and the aggregate price level fell in period 1 and consequently, the average firm believes that an aggregate productivity shock has occurred (although it cannot dismiss the possibility of a monetary shock being the source of the increased real GDP). As such, the simple-average firm also reduces its price, leading to a larger movement in real GDP. Over time, firms quite quickly disregard the likelihood of a monetary policy shock being the cause but only reduce their expectations of an aggregate TFP shock gradually.

Under the baseline specification, movements in real GDP and the aggregate price

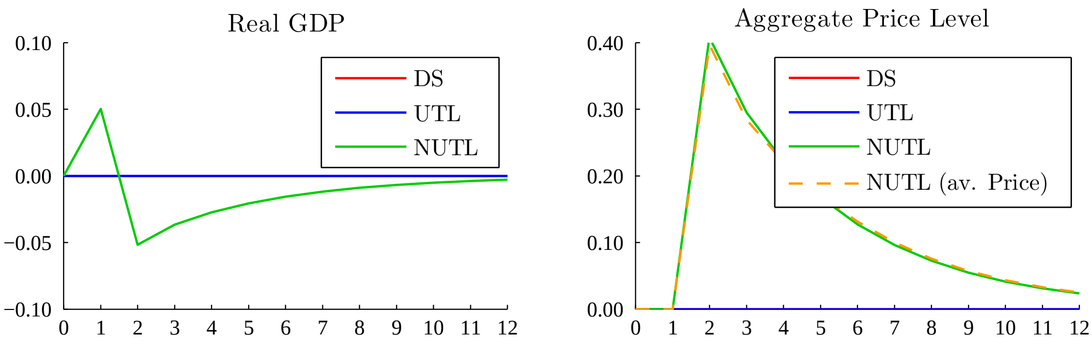


(b) Hierarchy of simple-average expectations under NUTL preferences

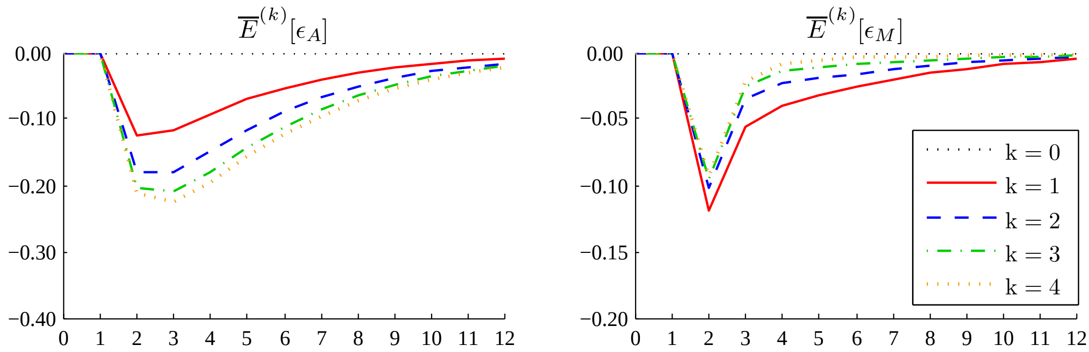
Figure 4.8: Responses to a 1 s.d. shock to $\widetilde{v_t^A}$

level are roughly half of the magnitude of those following a true aggregate productivity shock (it remains the case that the magnitude of the price level movement is quite a bit larger than that of real GDP) and somewhat less persistent. This stands in contrast to the previous chapter that examined the aggregate effect from network shocks, in which responses were roughly an order of magnitude smaller.

For a preference distribution shock that favours more expensive firms (figure 4.9), firms receive no signal of the shocks on impact and so do not adjust their prices, but real GDP increases mechanically. In period 2, seeing the increase in real GDP and no movement in the aggregate price from period 1, firms attribute the shock to both aggregate TFP and monetary policy. Both beliefs contribute to an increase in firms' prices in period 2, causing overall demand to fall enough to counteract the



(a) Aggregate IRFs



(b) Hierarchy of simple-average expectations under NUTL preferences

Figure 4.9: Responses to a 1 s.d. shock to $\widetilde{v_t^{\alpha}}$

mechanical increase in real GDP that emerges from the shock. The drop in real GDP and increase in the aggregate price in period 2 induces the average firm in period 3 to largely unwind its belief in a monetary policy shock, but remains consistent with an aggregate TFP shock. Firms' prices subsequently remain elevated and real GDP below trend until the shock dissipates.

Relative to an aggregate demand shock (to monetary policy), a shock to the distribution of demand under the baseline specificiation induces a price level response of roughly half the magnitude, but a relatively much smaller response in real GDP.

Variance decomposition

Table 4.2 shows the share of unconditional variance in real GDP and the aggregate price level that can be attributed to idiosyncratic shocks under NUTL preferences and the baseline parameterisation for different degrees of asymmetry in steady-state distribution of firms' base market shares. For Near-Uniform TransLog preferences

ζ^*	Real GDP	Aggregate price level
0	0	0
0.02	0.19	1.54
0.04	0.35	3.22
0.06	0.49	5.01
0.08	0.62	6.87
0.10	0.74	8.78

Table 4.2: Share of unconditional variance attributable to idiosyncratic shocks (%)

under the baseline parameterisation ($\zeta^* = 0.063$), roughly 0.5% of unconditional volatility in real GDP and 5.0% unconditional volatility in the aggregate price level are attributable to idiosyncratic shocks. Recall that the baseline parameterisation corresponds to a Zeta distribution with a shape parameter of $\gamma = 1.25$.

4.7 Conclusion

This chapter has discussed the possibility of making use of TransLog preferences in a macroeconomic model when household preferences are not uniform across goods and firms are free to adjust their prices every period (thereby avoiding both the Lucas critique under purely time-based price setting rules and the complexity of aggregation under state-based pricing systems) but face incomplete information.

Under TransLog preferences, firms continue to set their price as a mark-up over their nominal marginal costs, but the mark-up is endogenous. A price rise lowers a firm's share of household expenditure, which raises its own-price elasticity of demand and, hence, lowers its optimal mark-up, thereby dampening any response to a change in marginal cost. Under full information, we solve for an explicit solution to the firm's pricing problem and note that an endogenous market-exit condition arises in response to sufficiently large shocks to marginal cost (that necessitate prices sufficiently high as to cause the firm's expenditure share to be zero).

Under incomplete information, firms' signal extraction problem implies that (a) larger firms will place relatively more weight on movements in their marginal costs, while smaller firms will place more weight on movements in competitors' prices or consumer preferences; and (b) when considering competitors' prices, every firm must consider two distinct sums: the aggregate price level (as part of its estimation of nominal marginal cost) and a firm-specific weighted sum of competitors' prices.

The pricing rule under full TransLog preferences is, in general, intractable to aggregation. This is because each firm, in considering a weighted sum of other firms' prices, must also consider the weighted-average price specific to each of its competitors. This recursion creates an explosion in the size of the state vector quite apart from that arising from higher-order expectations (see section 1.2.1 in chapter 1). To simplify, we suppose that the representative household exhibits TransLog preferences with *near-uniformity in steady-state preferences*: firms' steady-state shares of household expenditure and their mark-ups are all equal, but their marginal costs and prices are not. With Near-Uniform TransLog (NUTL) preferences, all firms are interested in estimating the same two sums of individual prices: A weighted-average price[18] that comprises the aggregate price level and a simple-average price that affects movements in their mark-ups.

We compare the performance of NUTL preferences in a small DSGE model to corresponding models with fully Uniform TransLog (UTL) or Dixit-Stiglitz (DS) preferences. UTL and DS preferences are shown to induce very similar impulse responses following shocks to aggregate TFP or monetary policy. In contrast, impulse responses in the NUTL framework were generally more subdued, more hump-shaped and considerably more persistent. Despite firms experiencing no nominal rigidity and observing the same signals as under UTL or DS preferences, monetary shocks are clearly non-neutral beyond the period in which the shock occurs.

The inclusion of weighted-average prices also implies that the law of large numbers may not hold, so that firms' idiosyncratic shocks can have aggregate effects on the economy. In the context of NUTL preferences, this occurs when firms with large steady-state marginal costs experience shocks in one direction while firms with small steady-state marginal costs experience shocks in the other. We show that under parameterisations that match the distribution of firm size in the United States, idiosyncratic shocks in this sense may contribute 0.5% of unconditional volatility in real GDP and 5.0% of volatility in the price level.

[18] With the weights given by firms' steady-state base market share.

Appendix 4.A Proofs

This chapter contains derivations and proofs of results in the main paper.

4.A.1 Own-price super-elasticity of demand within the Almost Ideal Demand System

Taking note of equation (4.4):

$$\frac{P(i)\,Q(i)}{N} \equiv s(i) = \alpha(i) + \sum_j \gamma_{ij} \ln(P(j)) + \beta_i \ln\left(\frac{N}{\mathbb{P}}\right)$$

and equation (4.9):

$$\varepsilon_{ii} = 1 - \frac{1}{s(i)}\left[\gamma_{ii} - \beta_i\left(\alpha(i) + \sum_k \gamma_{ki} \ln(P(k))\right)\right]$$

The super-elasticity of demand is derived as follows, again taking income and other firms' prices as given:

$$\begin{aligned}
\xi_{ii} &\equiv \frac{\partial \ln \varepsilon_{ii}}{\partial \ln P(i)} \\
&= \frac{1}{\varepsilon_{ii}}\left[-\frac{\partial \frac{1}{s(i)}\left[\gamma_{ii} - \beta_i\left(\alpha(i) + \sum_k \gamma_{ki} \ln(P(k))\right)\right]}{\partial \ln P(i)}\right] \\
&= \frac{1}{\varepsilon_{ii}}\left[\frac{\beta_i \gamma_{ii}}{s(i)} + \frac{1}{s(i)}\left[\gamma_{ii} - \beta_i\left(\alpha(i) + \sum_k \gamma_{ki} \ln(P(k))\right)\right]\frac{1}{s(i)}\frac{\partial s(i)}{\partial \ln P(i)}\right] \\
&= \frac{1}{\varepsilon_{ii}}\left[\frac{\beta_i \gamma_{ii}}{s(i)} + \left\{\frac{1}{s(i)}\left[\gamma_{ii} - \beta_i\left(\alpha(i) + \sum_k \gamma_{ki} \ln(P(k))\right)\right]\right\}^2\right] \\
&= \frac{1}{\varepsilon_{ii}}\left[\frac{\beta_i \gamma_{ii}}{s(i)} + (\varepsilon_{ii} - 1)^2\right]
\end{aligned}$$

which is equation (4.17) in the main text.

4.A.2 Proof of proposition 3: Explicit solution for the one-period optimal price under full information

Defining lower-case letters to be the (natural) log of their upper-case counterparts – e.g. $x(i) \equiv \ln(X(i))$ – and substituting equation (4.14) in for $s(i)$, we can rewrite the monopolist's one-period profit (4.44) as

$$\Pi\left(i\right)=\left(\alpha\left(i\right)+\gamma_{ii}p\left(i\right)+\sum_{j\neq i}\gamma_{ij}p\left(j\right)\right)e^{n}-C\left(\left(\left(\alpha\left(i\right)+\gamma_{ii}p\left(i\right)+\sum_{j\neq i}\gamma_{ij}p\left(j\right)\right)e^{n-p(i)}\right)\right.$$

Taking all parameters, other firms' prices and aggregate expenditure as given, the optimal price for a monopolist when free to adjust their price every period is therefore found at the point $p\left(i\right)$ such that the following first-order condition is satisfied

$$\gamma_{ii}e^{n}-C'\left(Q\left(i\right)\right)\left(\gamma_{ii}e^{n-p(i)}-\left(\alpha\left(i\right)+\gamma_{ii}p\left(i\right)+\sum_{j\neq i}\gamma_{ij}p\left(j\right)\right)e^{n-p(i)}\right)=0$$

Defining $mc\left(i\right)\equiv\ln\left(C'\left(Q\left(i\right)\right)\right)$ as the log of marginal costs, we can rearrange this to give

$$\gamma_{ii}\left(p\left(i\right)+e^{p(i)-mc(i)}\right)=\gamma_{ii}-\alpha\left(i\right)-\sum_{j\neq i}\gamma_{ij}p\left(j\right)$$

or

$$p\left(i\right)+e^{p(i)-mc(i)}=\underbrace{1-\frac{\alpha\left(i\right)}{\gamma_{ii}}-\sum_{j\neq i}\frac{\gamma_{ij}}{\gamma_{ii}}p\left(j\right)}_{\equiv\phi(i)}$$

Denoting the right-hand side by $\phi\left(i\right)$, this in turn rearranges to

$$\ln\left(P\left(i\right)\right)+\frac{P\left(i\right)}{C'\left(Q\left(i\right)\right)}=\phi\left(i\right)\tag{4.72}$$

Taking the exponential of both sides and then dividing both sides by $C'\left(Q\left(i\right)\right)$ gives

$$\left(\frac{P\left(i\right)}{C'\left(Q\left(i\right)\right)}\right)e^{\left(\frac{P(i)^{*}}{C'(Q(i))}\right)}=\frac{1}{C'\left(Q\left(i\right)\right)}e^{\phi(i)}$$

The left-hand side is now in the form of the inverse of the Lambert $\mathcal{W}$ function, so we can therefore write

$$\frac{P\left(i\right)}{C'\left(Q\left(i\right)\right)}=\mathcal{W}\left(\frac{e^{\phi(i)}}{C'\left(Q\left(i\right)\right)}\right)\tag{4.73}$$

which is equation (4.46) in the main text.

First and Second Derivatives

$$
\begin{aligned}
\frac{\partial P(i)}{\partial MC(i)} &= \mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right) - \mathcal{W}'\left(\frac{e^{\phi(i)}}{MC(i)}\right)\frac{e^{\phi(i)}}{MC(i)} \\
&= \mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right) - \left(\frac{\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{\frac{e^{\phi(i)}}{MC(i)}\left(1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)\right)}\right)\frac{e^{\phi(i)}}{MC(i)} \\
&= \frac{\left[\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)\right]^2}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)} \\
&> 0
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 P(i)}{\partial MC(i)^2} &= 2\frac{\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}\mathcal{W}'\left(\frac{e^{\phi(i)}}{MC(i)}\right)\left(-\frac{e^{\phi(i)}}{[MC(i)]^2}\right) \\
&\quad - \left[\frac{\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}\right]^2\mathcal{W}'\left(\frac{e^{\phi(i)}}{MC(i)}\right)\left(-\frac{e^{\phi(i)}}{[MC(i)]^2}\right) \\
&= \left(2-\frac{\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}\right)\frac{\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}\mathcal{W}'\left(\frac{e^{\phi(i)}}{MC(i)}\right)\left(-\frac{e^{\phi(i)}}{[MC(i)]^2}\right) \\
&= -\left(\frac{2+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}\right)\frac{\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}\mathcal{W}'\left(\frac{e^{\phi(i)}}{MC(i)}\right)\frac{e^{\phi(i)}}{[MC(i)]^2} \\
&= -\left(\frac{2+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}\right)\left[\frac{\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}\right]^2\frac{1}{MC(i)} \\
&< 0
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial P(i)}{\partial P(j)} &= \mathcal{W}'\left(\frac{e^{\phi(i)}}{MC(i)}\right)e^{\phi(i)}\frac{\partial \phi(i)}{\partial P(j)} \\
&= \left(\frac{\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}\right)\frac{MC(i)}{P(j)}\left(-\frac{\gamma_{ij}}{\gamma_{ii}}\right) \\
&> 0
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 P(i)}{\partial P(j)^2} &= \frac{1}{\left[1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)\right]^2} \frac{MC(i)}{P(j)} \left(-\frac{\gamma_{ij}}{\gamma_{ii}}\right) \mathcal{W}'\left(\frac{e^{\phi(i)}}{MC(i)}\right) \frac{e^{\phi(i)}}{MC(i)} \frac{\partial \phi(i)}{\partial P(j)} \\
&\quad - \left(\frac{\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}\right) \frac{MC(i)}{[P(j)]^2} \left(-\frac{\gamma_{ij}}{\gamma_{ii}}\right) \\
&= \left(\frac{\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}\right) \frac{MC(i)}{[P(j)]^2} \left(-\frac{\gamma_{ij}}{\gamma_{ii}}\right) \left[\frac{1}{\left[1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)\right]^2} \left(-\frac{\gamma_{ij}}{\gamma_{ii}}\right) - 1\right] \\
&< 0
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial P(i)}{\partial \alpha(i)} &= \mathcal{W}'\left(\frac{e^{\phi(i)}}{MC(i)}\right) e^{\phi(i)} \frac{\partial \phi(i)}{\partial \alpha(i)} \\
&= \left(\frac{\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}\right) MC(i) \left(-\frac{1}{\gamma_{ii}}\right) \\
&> 0
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 P(i)}{\partial \alpha(i)^2} &= \frac{1}{\left[1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)\right]^2} MC(i) \left(-\frac{1}{\gamma_{ii}}\right) \mathcal{W}'\left(\frac{e^{\phi(i)}}{MC(i)}\right) \frac{e^{\phi(i)}}{MC(i)} \frac{\partial \phi(i)}{\partial \alpha(i)} \\
&= \frac{\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{\left[1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)\right]^3} \left(-\frac{1}{\gamma_{ii}}\right)^2 MC(i) \\
&> 0
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 P(i)}{\partial MC(i)\, \partial P(j)} &= \left(\frac{2+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}\right) \left(\frac{\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}\right) \mathcal{W}'\left(\frac{e^{\phi(i)}}{MC(i)}\right) \frac{e^{\phi(i)}}{MC(i)} \frac{\partial \phi(i)}{\partial P(j)} \\
&= \left(\frac{2+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}\right) \left[\frac{\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}\right]^2 \frac{1}{P(j)} \left(-\frac{\gamma_{ij}}{\gamma_{ii}}\right) \\
&> 0
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 P(i)}{\partial MC(i)\,\partial\alpha(i)} &= \left(\frac{2+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}\right)\left(\frac{\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}\right)\mathcal{W}'\left(\frac{e^{\phi(i)}}{MC(i)}\right)\frac{e^{\phi(i)}}{MC(i)}\frac{\partial\phi(i)}{\partial\alpha(i)} \\
&= \left(\frac{2+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}\right)\left[\frac{\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}\right]^2\left(-\frac{1}{\gamma_{ii}}\right) \\
&> 0
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 P(i)}{\partial P(j)\,\partial\alpha(i)} &= \frac{1}{\left[1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)\right]^2}\frac{MC(i)}{P(j)}\left(-\frac{\gamma_{ij}}{\gamma_{ii}}\right)\mathcal{W}'\left(\frac{e^{\phi(i)}}{MC(i)}\right)\frac{e^{\phi(i)}}{MC(i)}\frac{\partial\phi(i)}{\partial\alpha(i)} \\
&= \frac{1}{\left[1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)\right]^2}\left(\frac{\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}{1+\mathcal{W}\left(\frac{e^{\phi(i)}}{MC(i)}\right)}\right)\frac{MC(i)}{P(j)}\left(-\frac{\gamma_{ij}}{\gamma_{ii}}\right)\left(-\frac{1}{\gamma_{ii}}\right) \\
&> 0
\end{aligned}
$$

Existence and uniqueness of an equilibrium

The price-setting equation under full information (4.46) represents a mapping

$$T : \mathbb{R}_+^J \to \mathbb{R}_+^J \tag{4.74}$$

and a Nash equilibrium in prices will be a fixed point of this mapping. Clearly $\boldsymbol{P}_t = \mathbf{0}$ will be one such fixed point, but we dismiss this as trivial. In order to demonstrate the uniqueness of a non-zero vector of prices such that $\boldsymbol{P}_t^* = T(\boldsymbol{P}_t^*)$, note that it is insufficient to make use of the Banach (Contraction) Fixed-Point Theorem and Blackwell's sufficient conditions for demonstrating contraction (Blackwell, 1965), as T is not a contraction: the derivatives are unbounded as prices approach zero.

Instead, we note that as T is concave and satisfies the following:

$$
\begin{aligned}
T(\mathbf{0}) &= \mathbf{0} \\
\lim_{P(j)\to 0} P(i) &= \infty \;\forall j \neq i \\
\lim_{P(j)\to \infty} P(i) &= 0 \;\forall j \neq i
\end{aligned}
$$

then it meets the requirements laid out by Kennan (2001) for uniqueness. Because of its concavity and monotonicity, the mapping $T(\boldsymbol{P}_t)$ will therefore converge to the unique $\boldsymbol{P}_t^*$ for any strictly positive starting vector of prices.

4.A.3 Static pricing rule under incomplete information

As both settings are with static pricing, the first-order condition for the full-information case must hold in expectation here:

$$E_t(i)\left[\gamma_{ii}e^{n_t}\right] = E_t(i)\left[\gamma_{ii}e^{n_t+mc_t(i)-p_t(i)}\left(1-\frac{1}{\gamma_{ii}}\left(\alpha_t(i)+\gamma_{ii}p_t(i)+\sum_{j\neq i}\gamma_{ij}p_t(j)\right)\right)\right]$$

We suppose that the Γ matrix remains constant and known, while all other variables may vary over time. Denoting variables with an asterisk as being that variable in the no-shock, full-information equilibrium and variables with a hat above them to be that variable's deviation from the no-shock equilibrium – that is, $\widehat{p}_t(i) \equiv p_t(i) - p(i)^*$, etc. – a first-order Taylor series approximation of the left-hand side around the no-shock equilibrium is simply

$$LHS \approx \gamma_{ii}e^{n_t^*} + \gamma_{ii}e^{n_t^*}E(i)\left[\widehat{n}_t\right]$$

Next, note that

$$\left(1-\frac{1}{\gamma_{ii}}\left(\alpha_t(i)+\gamma_{ii}p_t(i)+\sum_{j\neq i}\gamma_{ij}p_t(j)\right)\right) = 1-\frac{s_t(i)}{\gamma_{ii}} = \mu_t(i)$$

so that a first-order approximation of the right-hand side is given by

$$\begin{aligned} RHS \approx\ & \gamma_{ii}e^{n_t^*+mc_t(i)^*-p_t(i)^*}\mu_t(i)^* \\ & + \gamma_{ii}e^{n_t^*+mc_t(i)^*-p_t(i)^*}\mu_t(i)^*E(i)\left[\widehat{n}_t+\widehat{mc}_t(i)-\widehat{p}_t(i)\right] \\ & - e^{n_t^*+mc_t(i)^*-p_t(i)^*}E(i)\left[\widehat{\alpha}_t(i)+\gamma_{ii}\widehat{p}_t(i)+\sum_{j\neq i}\gamma_{ij}\widehat{p}_t(j)\right] \end{aligned}$$

Equating these, noting that the first terms on each side must cancel out and rearranging slightly then gives

$$\begin{aligned} e^{p_t(i)^*-mc_t(i)^*}E(i)\left[\widehat{n}_t\right] = \ & \mu_t(i)^*E(i)\left[\widehat{n}_t+\widehat{mc}_t(i)-\widehat{p}_t(i)\right] \\ & - E(i)\left[\frac{1}{\gamma_{ii}}\widehat{\alpha}_t(i)+\widehat{p}_t(i)+\sum_{j\neq i}\frac{\gamma_{ij}}{\gamma_{ii}}\widehat{p}_t(j)\right] \end{aligned}$$

Recognising that $e^{p_t(i)^* - mc_t(i)^*} = \mu_t(i)^*$, we see that the terms in $\widehat{n}_t$ cancel out and we arrive, finally, at

$$\widehat{p}_t(i) = \frac{\mu_t(i)^*}{1+\mu_t(i)^*} E_t(i)\left[\widehat{mc}_t(i)\right] - \left(\frac{1}{1+\mu_t(i)^*}\right) E_t(i)\left[\frac{1}{\gamma_{ii}}\widehat{\alpha}_t(i) + \sum_{j\neq i}\frac{\gamma_{ij}}{\gamma_{ii}}\widehat{p}_t(j)\right]$$

which is equation (4.52) in the main text.

The first-order approximation of firm i's nominal marginal cost(4.30) around steady-state is:

$$\begin{aligned}\widehat{mc}_t(i) = {} & \widehat{w}_t - (1+\eta)\widehat{a}_t(i) + \eta\left(\widehat{y}_t + \widehat{\mathbb{p}}_t - \widehat{p}_t(i)\right) \\ & + \frac{\eta}{s(i)^*}\left(\widehat{\alpha}_t(i) + \gamma_{ii}\widehat{p}_t(i) + \textstyle\sum_{j\neq i}\gamma_{ij}\widehat{p}_t(j)\right)\end{aligned} \tag{4.75}$$

Recall that $\widehat{w}_t$ is the (log deviation in the) nominal wage. We can replace it by $\widehat{w}_t \equiv \widehat{\varpi}_t + \widehat{\mathbb{p}}_t$, where $\widehat{\varpi}_t$ is the real wage and $\widehat{\mathbb{p}}_t$ is the aggregate price level. Substituting this expression in for $\widehat{mc}_t(i)$ gives us

$$\begin{aligned}\widehat{p}_t(i) = {} & \left(\frac{\mu_t(i)^*}{1+\mu_t(i)^*}\right) E_t(i)\left[\begin{array}{c}\widehat{\varpi}_t + \widehat{\mathbb{p}}_t - (1+\eta)\widehat{a}_t(i) + \eta\left(\widehat{y}_t + \widehat{\mathbb{p}}_t - \widehat{p}_t(i)\right) \\ +\frac{\eta}{s(i)^*}\left(\widehat{\alpha}_t(i) + \gamma_{ii}\widehat{p}_t(i) + \sum_{j\neq i}\gamma_{ij}\widehat{p}_t(j)\right)\end{array}\right] \\ & + \left(\frac{1}{1+\mu_t(i)^*}\right) E_t(i)\left[\left(\frac{-1}{\gamma_{ii}}\right)\widehat{\alpha}_t(i) + \sum_{j\neq i}\left(\frac{-\gamma_{ij}}{\gamma_{ii}}\right)\widehat{p}_t(j)\right]\end{aligned}$$

Gathering like terms produces

$$\begin{aligned}\widehat{p}_t(i) = {} & \left(\frac{\mu_i^*(1+\eta)}{1+\mu_i^*(1+\eta\varepsilon_{ii}^*)}\right) E_t(i)\left[\widehat{\mathbb{p}}_t\right] \\ & + \left(\frac{\mu_i^*}{1+\mu_i^*(1+\eta\varepsilon_{ii}^*)}\right) E_t(i)\left[\widehat{\varpi}_t - (1+\eta)\widehat{a}_t(i) + \eta\widehat{y}_t\right] \\ & + \left(\frac{1+\eta\varepsilon_{ii}^*}{1+\mu_i^*(1+\eta\varepsilon_{ii}^*)}\right) E_t(i)\left[\left(\frac{-1}{\gamma_{ii}}\right)\widehat{\alpha}_t(i) + \sum_{j\neq i}\left(\frac{-\gamma_{ij}}{\gamma_{ii}}\right)\widehat{p}_t(j)\right]\end{aligned}$$

which, noting that $\frac{1+\eta\varepsilon_{ii}^*}{1+\mu_i^*\left(1+\eta\varepsilon_{ii}^*\right)} = 1 - \frac{\mu_i^*(1+\eta)}{1+\mu_i^*\left(1+\eta\varepsilon_{ii}^*\right)}$, is equation (4.53) in the main text.

4.A.4 Aggregation under near-uniformity in preferences

There are three aggregations to consider in the firm's pricing problem, two common and one firm-specific.

First, we have that individual firms are interested in firm-specific weighted averages of their competitors' prices

$$\sum_{j\neq i} \frac{-\gamma_{ij}}{\gamma_{ii}} p_t\left(j\right)$$

The assumption of uniformity-in-Γ (4.40a) turns this into

$$\sum_{j\neq i} \frac{-\gamma_{ij}}{\gamma_{ii}} p_t\left(j\right) = \frac{\gamma}{J}\frac{1}{J-1}\sum_{j\neq i} p_t\left(j\right)$$

Assumption (4.40b) then (a) supposes that each firm is sufficiently small that this expression is the same for everyone

$$\sum_{j\neq i} \frac{-\gamma_{ij}}{\gamma_{ii}} p_t\left(j\right) = \frac{\gamma}{J}\frac{1}{J-1}\sum_{j\neq i} p_t\left(j\right) = \frac{\gamma}{J}\overline{p_t}\ \forall i$$

and (b) applies a normalisation that when in steady-state, this simple average of log prices is zero (roughly, although not exactly, that the average steady-state price is unitary)

$$\sum_{j\neq i} \frac{-\gamma_{ij}}{\gamma_{ii}} p\left(j\right)^* = \frac{\gamma}{J}\frac{1}{J-1}\sum_{j\neq i} p\left(j\right)^* = \frac{\gamma}{J}\overline{p^*} = 0\ \forall i$$

Outside of steady-state, firms will therefore wish to estimate (note the hats to indicate deviations from steady-state)

$$\sum_{j\neq i} \frac{-\gamma_{ij}}{\gamma_{ii}} \widehat{p}_t\left(j\right) = \frac{\gamma}{J}\overline{\widehat{p}_t}\ \forall i$$

Second, the linearised aggregate price level (4.5) is

$$\widehat{\mathbb{p}}_t = \sum_i p\left(i\right)^* \widehat{\alpha}_t\left(i\right) + \sum_i \left(\alpha\left(i\right)^* + 2\sum_{j\neq i}\gamma_{ij} p\left(j\right)^*\right)\widehat{p}_t\left(i\right)$$

With uniformity-in-Γ (4.40a), this becomes

$$\widehat{\mathbb{p}}_t = \sum_i p\left(i\right)^* \widehat{\alpha}_t\left(i\right) + \sum_i \frac{1}{J}\left(\zeta_i + 2\gamma\overline{p^*}\right)\widehat{p}_t\left(i\right)$$

and with the normalisation that $\overline{p^*} = 0$, we get

$$\widehat{\mathbb{p}}_t = \sum_i p(i)^* \widehat{\alpha}_t(i) + \sum_i \alpha(i)^* \widehat{p}_t(i)$$

For the sum over prices, it is clear that since $\alpha(i)^* \in (0,1)$ and $\sum_i \alpha(i)^* = 1$, this is a weighted average all firms' prices, which we denote $\widetilde{\widehat{p}}_t$. Next, substitute in the expression for $p(i)^*$ – see equation (4.41) in the main text – and noting that $\widehat{\alpha}_t(i) = \frac{1}{J} v_t^\alpha$ – see equation (4.25a) – to get

$$\begin{aligned}\widehat{\mathbb{p}}_t &= \widetilde{\widehat{p}}_t + \sum_i \frac{1}{\gamma}(\zeta_i - 1)\frac{1}{J} v_t^\alpha(i) \\ &= \widetilde{\widehat{p}}_t + \frac{1}{\gamma}\sum_i \frac{1}{J}\zeta_i v_t^\alpha(i) - \frac{1}{\gamma}\frac{1}{J}\sum_i v_t^\alpha(i)\end{aligned}$$

The last term clearly goes to zero as $J \to \infty$, since it is a simple average and $E[v_t^\alpha(i)] = 0$. We are therefore left with

$$\widehat{\mathbb{p}}_t = \widetilde{\widehat{p}}_t + \frac{1}{\gamma}\widetilde{v}_t^\alpha \tag{4.76}$$

which is equation (4.57) in the main text.

Third, aggregate TFP (4.33) linearises as

$$\widehat{z}_t = \int \widehat{a}_t(i) + \widehat{p}_t(i) - \widehat{\mathbb{p}}_t - \frac{1}{s(i)^*}\left[\widehat{\alpha}_t(i) + \gamma_{ii}\widehat{p}_t(i) + \sum\nolimits_{j\neq i} \gamma_{ij}\widehat{p}_t(j)\right] di$$

Applying the uniformity-in-Γ (4.40a) assumption and using the definition of $\alpha_t(i)$, this becomes

$$\widehat{z}_t = \int \widehat{a}_t(i) + \widehat{p}_t(i) - \widehat{\mathbb{p}}_t - \frac{1}{s(i)^*}\left[\frac{1}{J}v_t^\alpha(i) - \frac{\gamma}{J}\widehat{p}_t(i) + \frac{\gamma}{J(J-1)}\sum\nolimits_{j\neq i}\widehat{p}_t(j)\right] di$$

Applying uniformity in steady-state expenditure shares (4.40c) and expanding $\widehat{a}_t(i)$, we get

$$\widehat{z}_t = \int \epsilon_t^A + v_t^A(i) + \widehat{p}_t(i) - \widehat{\mathbb{p}}_t - v_t^\alpha(i) + \gamma\widehat{p}_t(i) - \gamma\frac{1}{(J-1)}\sum\nolimits_{j\neq i}\widehat{p}_t(j)\, di$$

But then the last two terms cancel out and and we arrive at

$$\widehat{z}_t = \epsilon_t^A + \overline{\widehat{p}_t} - \widehat{\mathbb{p}}_t \tag{4.77}$$

Finally, we can linearise the household's intratemporal first-order condition (4.22) and the aggregate production function (4.32) to express the real wage as

$$\widehat{\varpi}_t = \left(\frac{1}{\sigma} + \frac{1}{\psi}(1+\eta)\right)\widehat{y}_t - \frac{1}{\psi}(1+\eta)\widehat{z}_t$$

Combined with the earlier linearised expressions for aggregate TFP (4.77) and the aggregate price level (4.76), we can substitute this into equation (4.56) to give

$$\begin{aligned}\widehat{p}_t(i) = {} & \theta\left(1+\frac{1}{\psi}\right)E_t(i)\left[\widetilde{\widehat{p}}_t\right] + \left(1-\theta\left(1+\frac{1}{\psi}\right)\right)E_t(i)\left[\overline{\widehat{p}}_t\right] \\ & + E_t(i)\left[\theta\left(\frac{\frac{1}{\sigma}+\eta}{1+\eta}+\frac{1}{\psi}\right)\widehat{y}_t - \theta\left(1+\frac{1}{\psi}\right)\epsilon_t^A + \theta\left(1+\frac{1}{\psi}\right)\frac{1}{\gamma}\widetilde{v}_t^\alpha - \theta v_t^A(i)\right]\end{aligned}$$

or, since $E_t(i)\left[\widetilde{v}_t^\alpha\right] = 0$,

$$\begin{aligned}\widehat{p}_t(i) = {} & \theta\left(1+\frac{1}{\psi}\right)E_t(i)\left[\widetilde{\widehat{p}}_t\right] + \left(1-\theta\left(1+\frac{1}{\psi}\right)\right)E_t(i)\left[\overline{\widehat{p}}_t\right] \\ & + E_t(i)\left[\theta\left(\frac{\frac{1}{\sigma}+\eta}{1+\eta}+\frac{1}{\psi}\right)\widehat{y}_t - \theta\left(1+\frac{1}{\psi}\right)\epsilon_t^A - \theta v_t^A(i)\right]\end{aligned}$$

which is equation (4.58) in the main text.

4.A.5 Higher-order expectations

With a linearisation of the household Eular equation, we therefore have the following system of equations:

$$\widehat{y}_t = E_t\left[\widehat{y}_{t+1}\right] - \sigma\left(\widehat{i}_t - E_t\left[\widehat{\mathbb{p}}_{t+1} - \widehat{\mathbb{p}}_t\right]\right)$$

$$\widehat{i}_t = \kappa_y \widehat{y}_t + \kappa_\pi E_t\left[\widehat{\mathbb{p}}_{t+1} - \widehat{p}_t\right] + \epsilon_t^M$$

$$\widehat{\mathbb{p}}_t = \begin{cases} \widetilde{\widehat{p}}_t + \frac{1}{\gamma}\widetilde{v}_t^\alpha & \text{for NUTL} \\ \overline{\widehat{p}}_t & \text{for UTL and DS} \end{cases}$$

and our three possible price-setting rules:

$$\begin{aligned}
\widehat{p}_t(i)^{TL} &= \theta^* E_t(i)\left[\widehat{\mathbb{p}}_t\right] + (1-\theta^*) E_t(i)\left[\overline{\widehat{p}_t}\right] \\
&\quad + \theta E_t(i)\left[\left(\frac{\frac{1}{\sigma}+\eta}{1+\eta}+\frac{1}{\psi}\right)\widehat{y}_t - \left(1+\frac{1}{\psi}\right)\epsilon_t^A - v_t^A(i)\right] \\
\widehat{p}_t(i)^{STL} &= E_t(i)\left[\overline{\widehat{p}_t}\right] \\
&\quad + \theta E_t(i)\left[\left(\frac{\frac{1}{\sigma}+\eta}{1+\eta}+\frac{1}{\psi}\right)\widehat{y}_t - \left(1+\frac{1}{\psi}\right)\epsilon_t^A - v_t^A(i)\right] \\
\widehat{p}_t(i)^{DS} &= E_t(i)\left[\overline{\widehat{p}_t}\right] \\
&\quad + \frac{1}{1+\varepsilon\eta} E_t(i)\left[\left(\frac{\frac{1}{\sigma}+\eta}{1+\eta}+\frac{1}{\psi}\right)\widehat{y}_t - \left(1+\frac{1}{\psi}\right)\epsilon_t^A - v_t^A(i)\right]
\end{aligned}$$

Substituting the first three expressions into the individual price equations, recognising that $E_t(i)\left[\widetilde{v}_t^{\alpha}\right] = 0$ and noting that firms observe $\widehat{a}_t(i)$ directly, we have

$$\begin{aligned}
\widehat{p}_t(i)^* = \lambda_1^* E_t(i)\left[\overline{\widehat{p}_t}\right] &+ \lambda_2^* E_t(i)\left[\widetilde{\widehat{p}_t}\right] \\
+ \lambda_3^* E_t(i)\left[\widehat{\mathbb{p}}_{t+1}\right] &+ \lambda_4^* E_t(i)\left[\widehat{y}_{t+1}\right] \\
+ \lambda_5^* E_t(i)\left[\epsilon_t^A\right] &+ \lambda_6^* E_t(i)\left[\epsilon_t^M\right] \\
+ \lambda_7^* \widehat{a}_t(i) &
\end{aligned}$$

for $* \in \{NUTL, UTL, DS\}$ and where the λ_*^* coefficients are given by:

Near-Uniform TransLog

$$\lambda_1^{NUTL} = 1-\theta\left(1+\frac{1}{\psi}\right)$$

$$\lambda_2^{NUTL} = \theta\left(1+\frac{1}{\psi}\right) - \theta\left(\frac{\frac{1}{\sigma}+\eta}{1+\eta}+\frac{1}{\psi}\right)\frac{\sigma}{1+\sigma\kappa_y}(1-\kappa_\pi)$$

$$\lambda_3^{NUTL} = \theta\left(\frac{\frac{1}{\sigma}+\eta}{1+\eta}+\frac{1}{\psi}\right)\frac{\sigma}{1+\sigma\kappa_y}(1-\kappa_\pi)$$

$$\lambda_4^{NUTL} = \theta\left(\frac{\frac{1}{\sigma}+\eta}{1+\eta}+\frac{1}{\psi}\right)\frac{1}{1+\sigma\kappa_y}$$

$$\lambda_5^{NUTL} = -\theta\left(1+\frac{1}{\psi}\right)$$

$$\lambda_6^{NUTL} = -\theta\left(\frac{\frac{1}{\sigma}+\eta}{1+\eta}+\frac{1}{\psi}\right)\frac{\sigma}{1+\sigma\kappa_y}$$

$$\lambda_7^{NUTL} = -\theta$$

Uniform TransLog

$$\lambda_1^{UTL} = 1 - \theta \left(\frac{\frac{1}{\sigma} + \eta}{1+\eta} + \frac{1}{\psi} \right) \frac{\sigma}{1+\sigma\kappa_y} (1-\kappa_\pi) \quad \lambda_2^{UTL} = 0$$

$$\lambda_3^{UTL} = \theta \left(\frac{\frac{1}{\sigma} + \eta}{1+\eta} + \frac{1}{\psi} \right) \frac{\sigma}{1+\sigma\kappa_y} (1-\kappa_\pi) \qquad \lambda_4^{UTL} = \theta \left(\frac{\frac{1}{\sigma} + \eta}{1+\eta} + \frac{1}{\psi} \right) \frac{1}{1+\sigma\kappa_y}$$

$$\lambda_5^{UTL} = -\theta \left(1 + \frac{1}{\psi} \right) \qquad \lambda_6^{UTL} = -\theta \left(\frac{\frac{1}{\sigma} + \eta}{1+\eta} + \frac{1}{\psi} \right) \frac{\sigma}{1+\sigma\kappa_y}$$

$$\lambda_7^{UTL} = -\theta$$

Dixit-Stiglitz

$$\lambda_1^{DS} = \xi \qquad \lambda_2^{DS} = 0$$

$$\lambda_3^{DS} = \frac{1+\eta}{1+\varepsilon\eta} \left(\frac{\frac{1}{\sigma} + \eta}{1+\eta} + \frac{1}{\psi} \right) \frac{\sigma}{1+\sigma\kappa_y} (1-\kappa_\pi) \quad \lambda_4^{DS} = \frac{1+\eta}{1+\varepsilon\eta} \left(\frac{\frac{1}{\sigma} + \eta}{1+\eta} + \frac{1}{\psi} \right) \frac{1}{1+\sigma\kappa_y}$$

$$\lambda_5^{DS} = -\frac{1+\eta}{1+\varepsilon\eta} \left(1 + \frac{1}{\psi} \right) \qquad \lambda_6^{DS} = -\frac{1+\eta}{1+\varepsilon\eta} \left(\frac{\frac{1}{\sigma} + \eta}{1+\eta} + \frac{1}{\psi} \right) \frac{\sigma}{1+\sigma\kappa_y}$$

$$\lambda_7^{DS} = -\frac{1+\eta}{1+\varepsilon\eta}$$

where $\xi \equiv 1 - \frac{1+\eta}{1+\varepsilon\eta} \left(\frac{\frac{1}{\sigma}+\eta}{1+\eta} + \frac{1}{\psi} \right) \frac{\sigma}{1+\sigma\kappa_y} (1-\kappa_\pi)$ is the same as in the previous chapter.

In order to ensure stability in the system, we require that decreasing weight be applied to higher-order expectations. In other words, we require that

$$\lambda_1 \in (0,1)\,;\ \lambda_2 \in (0,1)\,;\ \text{and}\ \lambda_1 + \lambda_2 \in (0,1)$$

The first of these requires that ψ be sufficiently large to ensure that $\theta \left(1 + \frac{1}{\psi} \right) < 1$. For any of the parameterisations chosen here, this requirement is satisfied. Given the definitions of λ_1 and λ_2, the third will automatically be satisfied so long as the second is, while the second requires that

$$0 \leq \theta \left(1 + \frac{1}{\psi} \right) - \theta \left(\frac{\frac{1}{\sigma} + \eta}{1+\eta} + \frac{1}{\psi} \right) \frac{\sigma}{1+\sigma\kappa_y} (1-\kappa_\pi) \leq 1$$

Taking the simple average of the pricing equation (4.60), we have

$$\begin{aligned} \overline{\widehat{p}_t} &= \lambda_1 \overline{E}_t \left[\overline{\widehat{p}_t} \right] \\ &+ \overline{E}_t \left[\lambda_2 \widetilde{\widehat{p}_t} + \lambda_3 \widehat{\mathbb{p}}_{t+1} + \lambda_4 \widehat{y}_{t+1} + \lambda_5 \epsilon_t^A + \lambda_6 \epsilon_t^M \right] + \lambda_7 \epsilon_t^A \end{aligned}$$

Repeatedly substituting this back into itself gives

$$\overline{\widehat{p}}_t = \sum_{k=0}^{\infty} \lambda_1^k \left(\overline{E}_t^{(k)} \left[\lambda_7 \epsilon_t^A + \overline{E}_t \left[\lambda_2 \widetilde{\widehat{p}}_t + \lambda_3 \widehat{\mathbb{p}}_{t+1} + \lambda_4 \widehat{y}_{t+1} + \lambda_5 \epsilon_t^A + \lambda_6 \epsilon_t^M \right] \right] \right)$$

or

$$\overline{\widehat{p}}_t = \lambda_7 \epsilon_t^A + \sum_{k=0}^{\infty} \lambda_1^k \overline{E}_t^{(k+1)} \left[\lambda_2 \widetilde{\widehat{p}}_t + \lambda_3 \widehat{\mathbb{p}}_{t+1} + \lambda_4 \widehat{y}_{t+1} + (\lambda_5 + \lambda_1 \lambda_7) \epsilon_t^A + \lambda_6 \epsilon_t^M \right]$$

Putting this into (4.60) in place of $\overline{\widehat{p}}_t$, we have

$$\begin{aligned}
\widehat{p}_t(i) = {} & \lambda_1 E_t(i) \left[\begin{array}{l} \lambda_7 \epsilon_t^A \\ + \sum_{k=0}^{\infty} \lambda_1^k \overline{E}_t^{(k+1)} \left[\lambda_2 \widetilde{\widehat{p}}_t + \lambda_3 \widehat{\mathbb{p}}_{t+1} + \lambda_4 \widehat{y}_{t+1} + (\lambda_5 + \lambda_1 \lambda_7) \epsilon_t^A + \lambda_6 \epsilon_t^M \right] \end{array} \right] \\
& + \lambda_2 E_t(i) \left[\widetilde{\widehat{p}}_t \right] \\
& + \lambda_3 E_t(i) \left[\widehat{\mathbb{p}}_{t+1} \right] \quad + \lambda_4 E_t(i) \left[\widehat{y}_{t+1} \right] \\
& + \lambda_5 E_t(i) \left[\epsilon_t^A \right] \quad + \lambda_6 E_t(i) \left[\epsilon_t^M \right] \\
& + \lambda_7 \widehat{a}_t(i)
\end{aligned}$$

or, rearranging,

$$\begin{aligned}
\widehat{p}_t(i) = {} & E_t(i) \left[\sum_{k=0}^{\infty} \lambda_1^k \overline{E}_t^{(k)} \left[\lambda_2 \widetilde{\widehat{p}}_t + \lambda_3 \widehat{\mathbb{p}}_{t+1} + \lambda_4 \widehat{y}_{t+1} + (\lambda_5 + \lambda_1 \lambda_7) \epsilon_t^A + \lambda_6 \epsilon_t^M \right] \right] \\
& + \lambda_7 \widehat{a}_t(i)
\end{aligned}$$

which is equation (4.61) in the main text. Now taking a weighted average of this (using $\alpha(i)^*$ as the weights), we obtain

$$\begin{aligned}
\widetilde{\widehat{p}}_t = {} & \lambda_7 \left(\epsilon_t^A + \widetilde{v}_t \right) \\
& + \widetilde{E}_t \left[\sum_{k=0}^{\infty} \lambda_1^k \overline{E}_t^{(k)} \left[\lambda_2 \widetilde{\widehat{p}}_t + \lambda_3 \widehat{\mathbb{p}}_{t+1} + \lambda_4 \widehat{y}_{t+1} + (\lambda_5 + \lambda_1 \lambda_7) \epsilon_t^A + \lambda_6 \epsilon_t^M \right] \right]
\end{aligned}$$

Substituting this back into itself gives

$$\begin{aligned}
\widetilde{\widehat{p}}_t = {} & \lambda_7 \left(\epsilon_t^A + \widetilde{v}_t \right) \\
& + \widetilde{E}_t \left[\sum_{k=0}^{\infty} \lambda_1^k \overline{E}_t^{(k)} \left[\begin{array}{l} \lambda_2 \left(\begin{array}{l} \lambda_7 \left(\epsilon_t^A + \widetilde{v}_t \right) \\ + \widetilde{E}_t \left[\sum_{k=0}^{\infty} \lambda_1^k \overline{E}_t^{(k)} \left[\begin{array}{l} \lambda_2 \widetilde{\widehat{p}}_t \\ + \lambda_3 \widehat{\mathbb{p}}_{t+1} + \lambda_4 \widehat{y}_{t+1} \\ + (\lambda_5 + \lambda_1 \lambda_7) \epsilon_t^A + \lambda_6 \epsilon_t^M \end{array} \right] \right] \end{array} \right) \\ + \lambda_3 \widehat{\mathbb{p}}_{t+1} + \lambda_4 \widehat{y}_{t+1} \\ + (\lambda_5 + \lambda_1 \lambda_7) \epsilon_t^A + \lambda_6 \epsilon_t^M \end{array} \right] \right]
\end{aligned}$$

Repeating this substitution eventually leads to

$$\begin{aligned}\widetilde{\widehat{p}}_t &= \lambda_7\left(\epsilon_t^A+\widetilde{v}_t^A\right)\\ &+\widetilde{E}_t\left[\boldsymbol{\phi}'\mathbb{E}_t^{(0:\infty)}\left[\lambda_3\widehat{\mathbb{p}}_{t+1}+\lambda_4\widehat{y}_{t+1}+\left(\lambda_5+\left(\lambda_1+\lambda_2\right)\lambda_7\right)\epsilon_t^A+\lambda_6\epsilon_t^M+\lambda_2\lambda_7\widetilde{v}_t^A\right]\right]\end{aligned}$$

where $\mathbb{E}_t^{(0:\infty)}[\cdot]$ is the hierarchy of all permutations of the two compound expectations and $\boldsymbol{\phi}$ assigns geometrically decreasing weights to each, with λ_1 raised to the power of the number of orders of $\overline{E}_t[\cdot]$ and λ_2 raised to the power of the number of orders of $\widetilde{E}_t[\cdot]$.

For Near-Uniform TransLog preferences, we can plug this into equation (4.76) to obtain

$$\begin{aligned}\widehat{\mathbb{p}}_t^{NUTL} &= \lambda_7\left(\epsilon_t^A+\widetilde{v}_t^A\right)+\frac{1}{\gamma}\widetilde{v}_t^\alpha\\ &+\widetilde{E}_t\left[\boldsymbol{\phi}'\mathbb{E}_t^{(0:\infty)}\left[\lambda_3\widehat{\mathbb{p}}_{t+1}+\lambda_4\widehat{y}_{t+1}+\left(\lambda_5+\left(\lambda_1+\lambda_2\right)\lambda_7\right)\epsilon_t^A+\lambda_6\epsilon_t^M+\lambda_2\lambda_7\widetilde{v}_t^A\right]\right]\end{aligned}$$

which is equation (4.63) in the main text.

4.A.6 Firms' learning

Deriving the Kalman filter

The filter derived here closely follows that developed by Nimark (2011b) as a means of avoiding the doubling-up of the state vector more typical in the literature, thereby allowing more accurate simulation results when working with finite computing resources.

Denoting i's expectation formed with period-t information as $E_t(i)[\cdot]\equiv E[\cdot|\mathcal{I}_t(i)]$, our goal is to find a mean square error minimising[19] formula for $E_t(i)[X_t]$. To begin, we suppose (and verify below) that each firm's signal vector is given by

$$\boldsymbol{q}_t(i)=C_1X_t+C_2X_{t-1}+R_1\begin{bmatrix}v_t^A(i)\\ v_t^\alpha(i)\end{bmatrix}+R_2\begin{bmatrix}v_{t-1}^A(i)\\ v_{t-1}^\alpha(i)\end{bmatrix}+R_3\boldsymbol{e}_t+R_4\widetilde{\boldsymbol{v}}_{t-1}$$

[19] And hence, given that all shocks are mean zero, a variance-minimising estimator.

and that the innovation (i.e. the unexpected component) in the firm's signal vector may be written as:

$$\begin{aligned}\overrightarrow{\boldsymbol{q}}_{t|t-1}(i) &\equiv \boldsymbol{q}_t(i) - E_{t-1}(i)\left[\boldsymbol{q}_t(i)\right] \\ &= D\overrightarrow{X_{t-1|t-1}}(i) + C_1G_1\boldsymbol{u}_t + R_1\begin{bmatrix} v_t^A(i) \\ v_t^\alpha(i)\end{bmatrix} + R_2\begin{bmatrix} v_{t-1}^A(i) \\ v_{t-1}^\alpha(i)\end{bmatrix} \\ &+ (R_3 + C_1G_3)\,\boldsymbol{e}_t + (R_4 + C_1G_4)\,\widetilde{\boldsymbol{v}}_{t-1}\end{aligned}$$

where $D \equiv (C_1F + C_2)$ and $\overrightarrow{X_{t-1|t-1}}(i) \equiv X_{t-1} - E_{t-1}(i)\left[X_{t-1}\right]$ is i's contemporaneous error in estimating X_t. Derivations of the coefficients in these two expressions are provided below in appendix 4.A.7. Note that there is no term in $\widetilde{\boldsymbol{v}}_t$ here, as firms' only contemporaneous signal is their productivity shock, which does not rely on it. For the term in $\widetilde{\boldsymbol{v}}_{t-1}$, we note that firms cannot form a prior expectation (strictly, it is zero) because of the independence of idiosyncratic shocks and the assumption that we have a continuum of firms.

Since $\overrightarrow{\boldsymbol{q}}_t(i)$ contains only *new* information available in period t, it must be orthogonal to any of i's estimates based on information from earlier periods. We can therefore use the result that $E\left[w|y,z\right] = E\left[w|y\right] + E\left[w|z\right]$ when $y \perp z$, so that

$$\begin{aligned}E_t(i)\left[X_t\right] &= E\left[X_t|\mathcal{I}_{t-1}(i)\right] + E\left[X_t|\overrightarrow{\boldsymbol{q}}_t(i)\right] \\ &= E_{t-1}(i)\left[X_t\right] + K_t\overrightarrow{\boldsymbol{q}}_t(i)\end{aligned} \tag{4.78}$$

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. For this to be the best linear estimator, we require that K_t be such that $\overrightarrow{\boldsymbol{q}}_{t|t-1}(i)$ is orthogonal to the corresponding projection error, $X_t - K_t\overrightarrow{\boldsymbol{q}}_{t|t-1}(i)$. That is, we require that

$$E\left[\left(X_t - K_t\overrightarrow{\boldsymbol{q}}_{t|t-1}(i)\right)\overrightarrow{\boldsymbol{q}}_{t|t-1}(i)'\right] = 0 \tag{4.79}$$

Rearranging this gives

$$K_t = E\left[X_t\overrightarrow{\boldsymbol{q}}_{t|t-1}(i)'\right]\left(E\left[\overrightarrow{\boldsymbol{q}}_{t|t-1}(i)\,\overrightarrow{\boldsymbol{q}}_{t|t-1}(i)'\right]\right)^{-1} \tag{4.80}$$

Before evaluating this, note that we can rewrite the law of motion for the hidden state as

$$X_t = F\left(\overrightarrow{X_{t-1|t-1}}(i) + E_{t-1}(i)\left[X_{t-1}\right]\right) + G_1\boldsymbol{u}_t + G_2\widetilde{\boldsymbol{v}}_t + G_3\boldsymbol{e}_t + G_4\widetilde{\boldsymbol{v}}_{t-1}$$

The first term of equation (4.80) thus expands to:

$$E\left[X_t \overrightarrow{\boldsymbol{q}}_{t|t-1}(i)'\right] = E\left[\begin{array}{l} \left(\begin{array}{l} F\left(\overrightarrow{X_{t-1|t-1}}(i) + E_{t-1}(i)\left[X_{t-1}\right]\right) + G_1\boldsymbol{u}_t + G_2\widetilde{\boldsymbol{v}}_t \\ +G_3\boldsymbol{e}_t + G_4\widetilde{\boldsymbol{v}}_{t-1} \end{array}\right) \\ \times\left(\begin{array}{l} D\overrightarrow{X_{t-1|t-1}}(i) + C_1G\boldsymbol{u}_t + R_1\begin{bmatrix} v_t^A(i) \\ v_t^\alpha(i)\end{bmatrix} + R_2\begin{bmatrix} v_{t-1}^A(i) \\ v_{t-1}^\alpha(i)\end{bmatrix} \\ +\left(R_3 + C_1G_3\right)\boldsymbol{e}_t + \left(R_4 + C_1G_4\right)\widetilde{\boldsymbol{v}}_{t-1} \end{array}\right)' \end{array}\right]$$

which simplifies to

$$\begin{aligned} E\left[X_t \overrightarrow{\boldsymbol{q}}_{t|t-1}(i)'\right] &= FV_{t-1|t-1}D' + G_1\Sigma_{uu}G_1'C_1' \\ &\quad + G_3\Sigma_{ee}\left(R_3 + C_1G_3\right)' + G_4\Sigma_{\widetilde{v}\widetilde{v}}\left(R_4 + C_1G_4\right)' \end{aligned}$$

where $V_{t|t} \equiv E\left[\overrightarrow{X_{t|t}}(i)\,\overrightarrow{X_{t|t}}(i)'\right]$ is the variance-covariance matrix associated with $E_t(i)\left[X_t\right]$ and I have exploited the fact that $\alpha(i)^* \to 0$ as $J \to \infty$ to obtain zero covariance between individual idiosyncratic shocks and their aggregated counterparts. Given the symmetry of the problem across agents, although individual expectations may differ the variance of each estimate will be common. For the second term, we have that

$$\begin{aligned} &E\left[\overrightarrow{\boldsymbol{q}}_{t|t-1}(i)\,\overrightarrow{\boldsymbol{q}}_{t|t-1}(i)'\right] \\ &= E\left[\begin{array}{l} \left(\begin{array}{l} D\overrightarrow{X_{t-1|t-1}}(i) + C_1G\boldsymbol{u}_t + R_1\begin{bmatrix} v_t^A(i) \\ v_t^\alpha(i)\end{bmatrix} + R_2\begin{bmatrix} v_{t-1}^A(i) \\ v_{t-1}^\alpha(i)\end{bmatrix} \\ +\left(R_3 + C_1G_3\right)\boldsymbol{e}_t + \left(R_4 + C_1G_4\right)\widetilde{\boldsymbol{v}}_{t-1} \end{array}\right) \\ \times\left(\begin{array}{l} D\overrightarrow{X_{t-1|t-1}}(i) + C_1G\boldsymbol{u}_t + R_1\begin{bmatrix} v_t^A(i) \\ v_t^\alpha(i)\end{bmatrix} + R_2\begin{bmatrix} v_{t-1}^A(i) \\ v_{t-1}^\alpha(i)\end{bmatrix} \\ +\left(R_3 + C_1G_3\right)\boldsymbol{e}_t + \left(R_4 + C_1G_4\right)\widetilde{\boldsymbol{v}}_{t-1} \end{array}\right)' \end{array}\right] \end{aligned}$$

which simplifies to

$$\begin{aligned} E\left[\overrightarrow{\boldsymbol{q}}_{t|t-1}(i)\,\overrightarrow{\boldsymbol{q}}_{t|t-1}(i)'\right] &= DV_{t-1|t-1}D' + C_1G_1\Sigma_{uu}G_1'C_1' \\ &\quad + R_1\Sigma_{vv}R_1' + R_2\Sigma_{vv}R_2' \\ &\quad + \left(R_3 + C_1G_3\right)\Sigma_{ee}\left(R_3 + C_1G_3\right)' \\ &\quad + \left(R_4 + C_1G_4\right)\Sigma_{\widetilde{v}\widetilde{v}}\left(R_4 + C_1G_4\right)' \end{aligned}$$

where $\Sigma_{vv} = \left[\begin{smallmatrix} \sigma^2_{vA} & 0 \\ 0 & \sigma^2_{v\alpha} \end{smallmatrix}\right]$. All together, the Kalman gain is therefore given by

$$K_t = \begin{pmatrix} FV_{t-1|t-1}D' + G_1\Sigma_{uu}G_1'C_1' \\ +G_3\Sigma_{ee}\left(R_3 + C_1G_3\right)' + G_4\Sigma_{\widetilde{v}\widetilde{v}}\left(R_4 + C_1G_4\right)' \end{pmatrix} \tag{4.81}$$
$$\times \begin{bmatrix} DV_{t-1|t-1}D' + C_1G_1\Sigma_{uu}G_1'C_1' \\ +R_1\Sigma_{vv}R_1' + R_2\Sigma_{vv}R_2' \\ +\left(R_3 + C_1G_3\right)\Sigma_{ee}\left(R_3 + C_1G_3\right)' \\ +\left(R_4 + C_1G_4\right)\Sigma_{\widetilde{v}\widetilde{v}}\left(R_4 + C_1G_4\right)' \end{bmatrix}^{-1}$$

Evolution of the gain and variance matricies

First, since we can rewrite the state equation as

$$\begin{aligned} X_t - E_{t-1}(i)\left[X_t\right] &= FX_{t-1} + G_1\boldsymbol{u}_t + G_2\widetilde{\boldsymbol{v}}_t + G_3\boldsymbol{e}_t + G_4\widetilde{\boldsymbol{v}}_{t-1} - E_{t-1}(i)\left[X_t\right] \\ &= F\left(X_{t-1} - E_{t-1}(i)\left[X_{t-1}\right]\right) + G_1\boldsymbol{u}_t + G_2\widetilde{\boldsymbol{v}}_t + G_3\boldsymbol{e}_t + G_4\widetilde{\boldsymbol{v}}_{t-1} \end{aligned}$$

we have that

$$V_{t|t-1} = FV_{t-1|t-1}F' + G_1\Sigma_{uu}G_1' + G_2\Sigma_{\widetilde{v}\widetilde{v}}G_2' + G_3\Sigma_{ee}G_3' + G_4\Sigma_{\widetilde{v}\widetilde{v}}G_4' \tag{4.82}$$

Next, add X_t to each side of equation (4.78) and rearrange to get

$$X_t - E_t(i)\left[X_t\right] + K_t\overrightarrow{\boldsymbol{q}}_{t|t-1}(i) = X_t - E_{t-1}(i)\left[X_t\right]$$

Since the innovation is orthogonal to both the prior error, $X_t - E_{t-1}(i)\left[X_t\right]$, and the posterior error, $X_t - E_t(i)\left[X_t\right]$, the variance of the right-hand side must equal the sum of the variances on the left-hand side, so that

$$\begin{aligned} V_{t|t} &= V_{t|t-1} - K_tVar\left(\overrightarrow{\boldsymbol{q}}_{t|t-1}(i)\right)K_t' \\ &= V_{t|t-1} - K_tVar\begin{pmatrix} D\overrightarrow{X_{t-1|t-1}}(i) + C_1G\boldsymbol{u}_t + R_1\begin{bmatrix} v_t^A(i) \\ v_t^\alpha(i) \end{bmatrix} + R_2\begin{bmatrix} v_{t-1}^A(i) \\ v_{t-1}^\alpha(i) \end{bmatrix} \\ +\left(R_3 + C_1G_3\right)\boldsymbol{e}_t + \left(R_4 + C_1G_4\right)\widetilde{\boldsymbol{v}}_{t-1} \end{pmatrix}K_t' \\ &= V_{t|t-1} - K_t\begin{bmatrix} DV_{t-1|t-1}D' + C_1G_1\Sigma_{uu}G_1'C_1' \\ +R_1\Sigma_{vv}R_1' + R_2\Sigma_{vv}R_2' \\ +\left(R_3 + C_1G_3\right)\Sigma_{ee}\left(R_3 + C_1G_3\right)' \\ +\left(R_4 + C_1G_4\right)\Sigma_{\widetilde{v}\widetilde{v}}\left(R_4 + C_1G_4\right)' \end{bmatrix}K_t' \end{aligned} \tag{4.83}$$

Provided that F represents a contraction, then there will exist time-invariant Kalman gain and Variance matricies, found by iterating equations (4.81), (4.82) and (4.83) forward until convergence is achieved. The form of these matricies will be:

$$K = \begin{pmatrix} FVD' + G_1\Sigma_{uu}G_1'C_1' \\ +G_3\Sigma_{ee}\left(R_3 + C_1G_3\right)' + G_4\Sigma_{\widetilde{v}\widetilde{v}}\left(R_4 + C_1G_4\right)' \end{pmatrix} \times \begin{bmatrix} DVD' + C_1G_1\Sigma_{uu}G_1'C_1' \\ +R_1\Sigma_{vv}R_1' + R_2\Sigma_{vv}R_2' \\ +\left(R_3 + C_1G_3\right)\Sigma_{ee}\left(R_3 + C_1G_3\right)' \\ +\left(R_4 + C_1G_4\right)\Sigma_{\widetilde{v}\widetilde{v}}\left(R_4 + C_1G_4\right)' \end{bmatrix}^{-1}$$

$$V = F\left(V - K_t\begin{bmatrix} DVD' + C_1G_1\Sigma_{uu}G_1'C_1' \\ +R_1\Sigma_{vv}R_1' + R_2\Sigma_{vv}R_2' \\ +\left(R_3 + C_1G_3\right)\Sigma_{ee}\left(R_3 + C_1G_3\right)' \\ +\left(R_4 + C_1G_4\right)\Sigma_{\widetilde{v}\widetilde{v}}\left(R_4 + C_1G_4\right)' \end{bmatrix}K'\right)F' + G_1\Sigma_{uu}G_1' + G_2\Sigma_{\widetilde{v}\widetilde{v}}G_2' + G_3\Sigma_{ee}G_3' + G_4\Sigma_{\widetilde{v}\widetilde{v}}G_4'$$

The law of motion for the hierarchy of firms' expectations

First note that the recursive formulation of X_t

$$X_t \equiv \mathbb{E}_t^{(0:\infty)}\left[\boldsymbol{x}_t\right] = \begin{bmatrix} \boldsymbol{x}_t \\ \overline{E}_t\left[X_t\right] \\ \widetilde{E}_t\left[X_t\right] \end{bmatrix}$$

allows us to write

$$\overline{E}_t\left[X_t\right] = \begin{bmatrix} 0 & I & 0 \end{bmatrix} X_t = T_sX_t$$
$$\widetilde{E}_t\left[X_t\right] = \begin{bmatrix} 0 & 0 & I \end{bmatrix} X_t = T_wX_t$$

where T_s and T_w select the elements from X_t with simple- and weighted-average expectations as their outermost. Next, recalling that the innovation in the firm's signal vector is given by

$$\overrightarrow{\boldsymbol{q}}_t(i) = D\left(X_{t-1} - E_{t-1}(i)\left[X_{t-1}\right]\right) + C_1G_1\boldsymbol{u}_t + R_1\begin{bmatrix} v_t^A(i) \\ v_t^\alpha(i) \end{bmatrix} + R_2\begin{bmatrix} v_{t-1}^A(i) \\ v_{t-1}^\alpha(i) \end{bmatrix} + \left(R_3 + C_1G_3\right)\boldsymbol{e}_t + \left(R_4 + C_1G_4\right)\widetilde{\boldsymbol{v}}_{t-1}$$

we can take the simple average of the Kalman filter (4.78) to obtain

$$\begin{aligned}\overline{E}_t\left[X_t\right] = T_s X_t = \left\{F T_s + K D\left(I - T_s\right)\right\} X_{t-1} \\ + K C_1 G_1 \boldsymbol{u}_t \\ + K\left(R_3 + C_1 G_3\right)\boldsymbol{e}_t \\ + K\left(R_4 + C_1 G_4\right)\widetilde{\boldsymbol{v}}_{t-1}\end{aligned}$$

and the weighted average of the same to obtain

$$\begin{aligned}\widetilde{E}_t\left[X_t\right] = T_w X_t = \left\{F T_w + K D\left(I - T_w\right)\right\} X_{t-1} \\ + K C_1 G_1 \boldsymbol{u}_t \\ + K R_1 \widetilde{\boldsymbol{v}}_t \\ + K\left(R_3 + C_1 G_3\right)\boldsymbol{e}_t \\ + K\left(R_2 + R_4 + C_1 G_4\right)\widetilde{\boldsymbol{v}}_{t-1}\end{aligned}$$

We can therefore write

$$X_t = F X_{t-1} + G_1 \boldsymbol{u}_t + G_2 \widetilde{\boldsymbol{v}}_t + G_3 \boldsymbol{e}_t + G_4 \widetilde{\boldsymbol{v}}_{t-1}$$

where

$$F = \begin{bmatrix} \begin{bmatrix} \rho_\epsilon^A & 0 & 0_{1\times\infty} \\ 0 & \rho_\epsilon^M & 0_{1\times\infty} \end{bmatrix} \\ F T_s + K D\left(I - T_s\right) \\ F T_w + K D\left(I - T_w\right) \end{bmatrix} \quad G_1 = \begin{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ K C_1 G_1 \\ K C_1 G_1 \end{bmatrix}$$
$$G_2 = \begin{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\ 0_{\infty\times 2} \\ K R_1 \end{bmatrix} \quad G_3 = \begin{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\ K\left(R_3 + C_1 G_3\right) \\ K\left(R_3 + C_1 G_3\right) \end{bmatrix} \quad G_4 = \begin{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\ K\left(R_4 + C_1 G_4\right) \\ K\left(R_2 + R_4 + C_1 G_4\right) \end{bmatrix}$$

4.A.7 Solving the model

Given the law of motion for the hierarchy of firms' expectations:

$$X_t = F X_{t-1} + G_1 \boldsymbol{u}_t + G_2 \widetilde{\boldsymbol{v}}_t + G_3 \boldsymbol{e}_t + G_4 \widetilde{\boldsymbol{v}}_{t-1}$$

and the following expression for firms' signal vectors:

$$\boldsymbol{q}_t(i) = C_1 X_t + C_2 X_{t-1} + R_1 \begin{bmatrix} v_t^A(i) \\ v_t^\alpha(i) \end{bmatrix} + R_2 \begin{bmatrix} v_{t-1}^A(i) \\ v_{t-1}^\alpha(i) \end{bmatrix} + R_3 \boldsymbol{e}_t + R_4 \widetilde{\boldsymbol{v}}_{t-1}$$

we here demonstrate that the aggregate variables of the economy may then be characterised as:

$$\widehat{y}_t^* = \boldsymbol{\gamma}_y^{*\prime} X_t + \boldsymbol{\delta}_y^{*\prime} \widetilde{\boldsymbol{v}}_t$$
$$\widehat{\mathbb{p}}_t^* = \boldsymbol{\gamma}_p^{*\prime} X_t + \boldsymbol{\delta}_p^{*\prime} \widetilde{\boldsymbol{v}}_t$$

for $* \in \{NUTL, UTL, DS\}$.

Building an expression for $\widehat{y}_t$

We start with the expression for output in period t

$$\begin{aligned}(1+\sigma\kappa_y)\,\widehat{y}_t = E_t\left[\widehat{y}_{t+1}\right] \\ -\sigma(1-\kappa_\pi)\,\widehat{\mathbb{p}}_t \\ +\sigma(1-\kappa_\pi)\,E_t\left[\widehat{\mathbb{p}}_{t+1}\right] \\ -\sigma\epsilon_t^M\end{aligned}$$

Substituting in the conjectured solution and recalling that the representative household and central bank have full information then gives

$$\begin{aligned}(1+\sigma\kappa_y)\,\widehat{y}_t = \boldsymbol{\gamma}_y^{*\prime}\left(FX_t + G_4\widetilde{\boldsymbol{v}}_t\right) \\ -\sigma(1-\kappa_\pi)\,\boldsymbol{\gamma}_p^{*\prime} X_t \\ +\sigma(1-\kappa_\pi)\,\boldsymbol{\gamma}_p^{*\prime}\left(FX_t + G_4\widetilde{\boldsymbol{v}}_t\right) \\ -\sigma\epsilon_t^M\end{aligned}$$

Rearranging then gives

$$\widehat{y}_t = \boldsymbol{\gamma}_y^{*\prime} X_t + \boldsymbol{\delta}_y^{*\prime} \widetilde{\boldsymbol{v}}_t$$

where

$$\boldsymbol{\gamma}_y^{*\prime} = \frac{1}{1+\sigma\kappa_y}\left(\begin{array}{l} \boldsymbol{\gamma}_y^{*\prime} F \\ +\sigma(1-\kappa_\pi)\,\boldsymbol{\gamma}_p^{*\prime}(F-I) \\ +\begin{bmatrix} 0 & -\sigma & 0_{1\times\infty} \end{bmatrix} \end{array}\right)$$
$$\boldsymbol{\delta}_y^{*\prime} = \boldsymbol{\gamma}_y^{*\prime} G_4 + \sigma(1-\kappa_\pi)\,\boldsymbol{\gamma}_p^{*\prime} G_4$$

or

$$\boldsymbol{\gamma}_y^{*\prime} = \frac{1}{1+\sigma\kappa_y}\begin{pmatrix} \sigma\left(1-\kappa_\pi\right)\boldsymbol{\gamma}_p^{*\prime}\left(F-I\right) \\ +\begin{bmatrix}0 & -\sigma & 0_{1\times\infty}\end{bmatrix} \end{pmatrix}\left(I-\frac{1}{1+\sigma\kappa_y}F\right)^{-1}$$

Building an expression for $\widehat{\mathbb{p}}_t$ (Uniform TransLog and Dixit-Stiglitz preferences)

We start with the expression for the aggregate price:

$$\begin{aligned}\widehat{\mathbb{p}}_t^* &= \lambda_7^*\,\epsilon_t^A \\ &\quad+\sum_{k=0}^{\infty}\left(\lambda_1^*\right)^k\overline{E}_t^{(k+1)}\left[\lambda_3^*\widehat{\mathbb{p}}_{t+1}^*+\lambda_4^*\widehat{y}_{t+1}+\left(\lambda_5^*+\lambda_1^*\lambda_7^*\right)\epsilon_t^A+\lambda_6^*\,\epsilon_t^M\right]\end{aligned}$$

for $*\in\{UTL,DS\}$. Substituting in the conjectured solution and noting that firms' incomplete information implies that $E_t\left(i\right)\left[\widetilde{\boldsymbol{v}}_t\right]=0\ \forall i$, we have

$$\begin{aligned}\widehat{\mathbb{p}}_t^* &= \lambda_7^*\,\epsilon_t^A \\ &\quad+\sum_{k=0}^{\infty}\left(\lambda_1^*\right)^k\overline{E}_t^{(k+1)}\left[\lambda_3^*\boldsymbol{\gamma}_p^{*\prime}FX_t+\lambda_4^*\boldsymbol{\gamma}_y^{*\prime}FX_t+\left(\lambda_5^*+\lambda_1^*\lambda_7^*\right)\epsilon_t^A+\lambda_6^*\,\epsilon_t^M\right]\end{aligned}$$

or

$$\begin{aligned}\widehat{\mathbb{p}}_t^* &= \begin{bmatrix}\lambda_7^* & 0 & 0_{1\times\infty}\end{bmatrix}X_t \\ &\quad+\left(\lambda_3^*\boldsymbol{\gamma}_p^{*\prime}F+\lambda_4^*\boldsymbol{\gamma}_y^{*\prime}F+\begin{bmatrix}\left(\lambda_5^*+\lambda_1^*\lambda_7^*\right) & \lambda_6^* & 0_{1\times\infty}\end{bmatrix}\right)\sum_{k=0}^{\infty}\left(\lambda_1^*\right)^k\overline{E}_t^{(k+1)}\left[X_t\right]\end{aligned}$$

from which we can read that

$$\widehat{\mathbb{p}}_t^* = \boldsymbol{\gamma}_p^{*\prime}X_t$$

where

$$\begin{aligned}\widehat{\mathbb{p}}_t^* &= \begin{bmatrix}\lambda_7^* & 0 & 0_{1\times\infty}\end{bmatrix} \\ &\quad+\begin{pmatrix}\lambda_3^*\boldsymbol{\gamma}_p^{*\prime}F+\lambda_4^*\boldsymbol{\gamma}_y^{*\prime}F \\ +\begin{bmatrix}\left(\lambda_5^*+\lambda_1^*\lambda_7^*\right) & \lambda_6^* & 0_{1\times\infty}\end{bmatrix}\end{pmatrix}\left(I-\lambda_1^*T_s\right)^{-1}T_s\end{aligned}$$

Building an expression for $\widehat{\mathbb{p}}_t$ (Near-Uniform TransLog preferences)

We start with the expression for the aggregate price level in period t

$$\begin{aligned}\widehat{\mathbb{p}}_t^{NUTL} &= \lambda_7\left(\epsilon_t^A+\widetilde{v}_t^A\right)+\frac{1}{\gamma}\widetilde{v}_t^\alpha \\ &+\widetilde{E}_t\left[\boldsymbol{\phi}'\mathbb{E}_t^{(0:\infty)}\left[\lambda_3\widehat{\mathbb{p}}_{t+1}+\lambda_4\widehat{y}_{t+1}+\left(\lambda_5+\left(\lambda_1+\lambda_2\right)\lambda_7\right)\epsilon_t^A+\lambda_6\epsilon_t^M+\lambda_2\lambda_7\widetilde{v}_t^A\right]\right]\end{aligned}$$

Substituting in the conjectured solution and recalling that $E_t(i)\left[\widetilde{\boldsymbol{v}}_t\right]=\mathbf{0}\ \forall i$, we have

$$\begin{aligned}\widehat{\mathbb{p}}_t^{NUTL} &= \lambda_7\left(\epsilon_t^A+\widetilde{v}_t^A\right)+\frac{1}{\gamma}\widetilde{v}_t^\alpha \\ &+\widetilde{E}_t\left[\boldsymbol{\phi}'\mathbb{E}_t^{(0:\infty)}\left[\lambda_3\boldsymbol{\gamma}_p'FX_t+\lambda_4\boldsymbol{\gamma}_y'FX_t+\left(\lambda_5+\left(\lambda_1+\lambda_2\right)\lambda_7\right)\epsilon_t^A+\lambda_6\epsilon_t^M\right]\right]\end{aligned}$$

or

$$\begin{aligned}\widehat{\mathbb{p}}_t^{NUTL} &= \begin{bmatrix}\lambda_7 & 0 & 0_{1\times\infty}\end{bmatrix}X_t+\begin{bmatrix}\lambda_7 & \frac{1}{\gamma}\end{bmatrix}\widetilde{\boldsymbol{v}}_t \\ &+\begin{pmatrix}\left(\lambda_3\boldsymbol{\gamma}_p'+\lambda_4\boldsymbol{\gamma}_y'\right)F \\ \begin{bmatrix}\left(\lambda_5+\left(\lambda_1+\lambda_2\right)\lambda_7\right) & \lambda_6 & 0_{1\times\infty}\end{bmatrix}\end{pmatrix}\widetilde{E}_t\left[\boldsymbol{\phi}'\mathbb{E}_t^{(0:\infty)}\left[X_t\right]\right]\end{aligned}$$

To simplify this further, it will be necessary to consider the following object

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

Recalling the recursive formulation of $\mathbb{E}_t^{(0:\infty)}[\cdot]$, we have that

$$\begin{aligned}X_t &= \begin{bmatrix}\boldsymbol{x}_t \\ \overline{E}_t\left[X_t\right] \\ \widetilde{E}_t\left[X_t\right]\end{bmatrix} \\ \overline{E}_t\left[X_t\right] &= \begin{bmatrix}0 & I & 0\end{bmatrix}X_t=T_sX_t \\ \widetilde{E}_t\left[X_t\right] &= \begin{bmatrix}0 & 0 & I\end{bmatrix}X_t=T_wX_t\end{aligned}$$

where T_s and T_w select the elements from X_t with simple- and weighted-average expectations as their outermost. We can therefore write

$$\begin{aligned}\boldsymbol{\phi}'\mathbb{E}_t^{(0:\infty)}\left[X_t\right] = X_t & \\ &+\lambda_1\overline{E}_t\left[X_t\right]+\lambda_2\widetilde{E}_t\left[X_t\right] \\ &+\lambda_1^2\overline{E}_t\left[\overline{E}_t\left[X_t\right]\right]+\lambda_1\lambda_2\widetilde{E}_t\left[\overline{E}_t\left[X_t\right]\right] \\ &\quad+\lambda_1\lambda_2\overline{E}_t\left[\widetilde{E}_t\left[X_t\right]\right]+\lambda_2^2\widetilde{E}_t\left[\widetilde{E}_t\left[X_t\right]\right] \\ &+\cdots\end{aligned}$$

and then rewrite this as

$$\begin{aligned}\boldsymbol{\phi}'\mathbb{E}_t^{(0:\infty)}\left[X_t\right] = X_t& \\ &+ \lambda_1 T_s X_t + \lambda_2 T_2 X_t \\ &+ \lambda_1^2 T_s^2 X_t + \lambda_1\lambda_2 T_w T_s X_t \\ &\quad + \lambda_1\lambda_2 T_s T_w X_t + \lambda_2^2 T_w^2 X_t \\ &+ \cdots\end{aligned}$$

But this is then simply an infinite (and convergent) sum of binomials

$$\boldsymbol{\phi}'\mathbb{E}_t^{(0:\infty)}\left[X_t\right] = \left(\sum_{k=0}^{\infty}\left[\lambda_1 T_s + \lambda_2 T_w\right]^k\right) X_t$$

which may be written as

$$\boldsymbol{\phi}'\mathbb{E}_t^{(0:\infty)}\left[X_t\right] = \left(I - \left[\lambda_1 T_s + \lambda_2 T_w\right]\right)^{-1} X_t$$

Moving back to the expression for the aggregate price level, we can therefore write

$$\begin{aligned}\widehat{\mathbb{P}}_t^{NUTL} &= \begin{bmatrix}\lambda_7 & 0 & 0_{1\times\infty}\end{bmatrix} X_t + \begin{bmatrix}\lambda_7 & \frac{1}{\gamma}\end{bmatrix}\widetilde{\boldsymbol{v}}_t \\ &+ \begin{pmatrix}\left(\lambda_3\boldsymbol{\gamma}_p' + \lambda_4\boldsymbol{\gamma}_y'\right) F \\ \begin{bmatrix}\left(\lambda_5 + \left(\lambda_1 + \lambda_2\right)\lambda_7\right) & \lambda_6 & 0_{1\times\infty}\end{bmatrix}\end{pmatrix} \\ &\qquad \times \left(I - \left[\lambda_1 T_s + \lambda_2 T_w\right]\right)^{-1} T_w X_t\end{aligned}$$

or, gathering terms,

$$\widehat{\mathbb{P}}_t^{NUTL} = \left(\boldsymbol{\gamma}_p^{NUTL}\right)' X_t + \left(\boldsymbol{\delta}_p^{NUTL}\right)'\widetilde{\boldsymbol{v}}_t$$

where

$$\begin{aligned}\left(\boldsymbol{\gamma}_p^{NUTL}\right)' &= \begin{bmatrix}\lambda_7 & 0_{1\times\infty}\end{bmatrix} \\ &\quad + \begin{pmatrix}\left(\lambda_3\left(\boldsymbol{\gamma}_p^{NUTL}\right)' + \lambda_4\left(\boldsymbol{\gamma}_y^{NUTL}\right)'\right) F \\ \begin{bmatrix}\left(\lambda_5 + \left(\lambda_1 + \lambda_2\right)\lambda_7\right) & 0_{1\times\infty}\end{bmatrix}\end{pmatrix} \\ &\qquad \times \left(I - \left[\lambda_1 T_s + \lambda_2 T_w\right]\right)^{-1} T_w \\ \left(\boldsymbol{\delta}_p^{NUTL}\right)' &= \begin{bmatrix}\lambda_7 & \frac{1}{\gamma}\end{bmatrix}\end{aligned}$$

The firms' signal vector

Recall the signal vector observed each period

$$\boldsymbol{q}_t(i) = \begin{bmatrix} \widehat{y}_{t-1} + e_{y,t} \\ \widehat{\mathbb{p}}_{t-1} + e_{p,t} \\ \widehat{y}_{t-1}(i) \\ \widehat{w}_{t-1} \\ \widehat{a}_t(i) \end{bmatrix}'$$

With the conjectured solution, we can immediately write that

$$\begin{aligned}
\widehat{y}_{t-1} &= \boldsymbol{\gamma}_y^{*\prime} X_{t-1} + \boldsymbol{\delta}_y^{*\prime} \widetilde{\boldsymbol{v}}_{t-1} \\
\widehat{\mathbb{p}}_{t-1} &= \boldsymbol{\gamma}_p^{*\prime} X_{t-1} + \boldsymbol{\delta}_p^{*\prime} \widetilde{\boldsymbol{v}}_{t-1} \\
\widehat{a}_t(i) &= \begin{bmatrix} 1 & 0 & 0_{1\times\infty} \end{bmatrix} X_t + \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} v_t^A(i) \\ v_t^\alpha(i) \end{bmatrix}
\end{aligned}$$

which leaves only $\widehat{y}_{t-1}(i)$ and $\widehat{w}_{t-1}$ to work out.

Making use of the previous-period quantity (TransLog and Symmetric TransLog preferences)

To make use of the quantity sold from the previous period, we note again the two definitions of expenditure share in the TransLog framework:

$$\frac{P_{t-1}(i)\, Y_{t-1}(i)}{\mathbb{P}_{t-1} Y_{t-1}} = s_{t-1}(i) = \alpha_{t-1}(i) + \gamma_{ii} p_{t-1}(i) + \sum_{j \neq i} \gamma_{ij} p_{t-1}(j)$$

Linearising and making use of our near-uniformity assumptions and the definition of $\alpha_t(i)$, this becomes

$$\frac{1}{J}\left(\widehat{p}_{t-1}(i) + \widehat{y}_{t-1}(i) - \widehat{\mathbb{p}}_{t-1} - \widehat{y}_{t-1}\right) = \widehat{s}_{t-1}(i) = \frac{1}{J} v_{t-1}^\alpha(i) - \frac{\gamma}{J} \widehat{p}_{t-1}(i) + \frac{\gamma}{J} \overline{\widehat{p}_{t-1}}$$

Further rearranging, we may write this as

$$\widehat{p}_{t-1}(i) + \widehat{y}_{t-1}(i) = \widehat{y}_{t-1} + \widehat{\mathbb{p}}_{t-1} + \gamma \overline{\widehat{p}_{t-1}} + v_{t-1}^\alpha(i)$$

The left-hand side of this expression is made up entirely of objects observed by the firm. Since the firm's price is necessarily a linear function of its period $t-1$

information set, observing $\widehat{y}_{t-1}(i)$ is informationally equivalent to observing $\widehat{p}_{t-1}(i)+\widehat{y}_{t-1}(i)$. To construct this, we start with the earlier expression for $\overline{\widehat{p}_t}$:

$$\overline{\widehat{p}_t} = \lambda_7 \epsilon_t^A + \sum_{k=0}^{\infty} \lambda_1^k \overline{E}_t^{(k+1)} \left[\lambda_2 \widetilde{\widehat{p}_t} + \lambda_3 \widehat{\mathbb{p}}_{t+1} + \lambda_4 \widehat{y}_{t+1} + (\lambda_5 + \lambda_1 \lambda_7)\, \epsilon_t^A + \lambda_6 \epsilon_t^M\right]$$

Substituting the definition of the aggregate price under Near-Uniform TransLog preferences gives

$$\overline{\widehat{p}_t} = \lambda_7 \epsilon_t^A + \sum_{k=0}^{\infty} \lambda_1^k \overline{E}_t^{(k+1)} \left[\lambda_2 \widehat{\mathbb{p}}_t - \frac{1}{\gamma} \widetilde{v}_t^{\alpha} + \lambda_3 \widehat{\mathbb{p}}_{t+1} + \lambda_4 \widehat{y}_{t+1} + (\lambda_5 + \lambda_1 \lambda_7)\, \epsilon_t^A + \lambda_6 \epsilon_t^M\right]$$

Substituting in the conjectured solution then gives

$$\overline{\widehat{p}_t} = \lambda_7 \epsilon_t^A + \sum_{k=0}^{\infty} \lambda_1^k \overline{E}_t^{(k+1)} \left[\lambda_2 \boldsymbol{\gamma}_p' X_t + \lambda_3 \boldsymbol{\gamma}_p' F X_t + \lambda_4 \boldsymbol{\gamma}_y' F X_t + (\lambda_5 + \lambda_1 \lambda_7)\, \epsilon_t^A + \lambda_6 \epsilon_t^M\right]$$

Gathering like terms,

$$\begin{aligned}\overline{\widehat{p}_t} &= \begin{bmatrix}\lambda_7 & 0 & 0_{1\times\infty}\end{bmatrix} X_t \\ &\quad + \begin{pmatrix} \lambda_2 \boldsymbol{\gamma}_p' + \lambda_3 \boldsymbol{\gamma}_p' F + \lambda_4 \boldsymbol{\gamma}_y' F \\ \begin{bmatrix}(\lambda_5 + \lambda_1 \lambda_7) & \lambda_6 & 0_{1\times\infty}\end{bmatrix} \end{pmatrix} \sum_{k=0}^{\infty} \lambda_1^k \overline{E}_t^{(k+1)} [X_t]\end{aligned}$$

or

$$\overline{\widehat{p}_t} = \boldsymbol{\gamma}_{\overline{p}}' X_t$$

where

$$\begin{aligned}\boldsymbol{\gamma}_{\overline{p}}' &= \begin{bmatrix}\lambda_7 & 0 & 0_{1\times\infty}\end{bmatrix} \\ &\quad + \begin{pmatrix} \lambda_2 \boldsymbol{\gamma}_p' + \lambda_3 \boldsymbol{\gamma}_p' F + \lambda_4 \boldsymbol{\gamma}_y' F \\ \begin{bmatrix}(\lambda_5 + \lambda_1 \lambda_7) & \lambda_6 & 0_{1\times\infty}\end{bmatrix} \end{pmatrix} (I - \lambda_1 T_s)^{-1} T_s\end{aligned}$$

which is to say

$$\widehat{p}_{t-1}(i) + \widehat{y}_{t-1}(i) = (\boldsymbol{\gamma}_y + \boldsymbol{\gamma}_p + \gamma \boldsymbol{\gamma}_{\overline{p}})' X_{t-1} + (\boldsymbol{\delta}_y + \boldsymbol{\delta}_p)' \widetilde{\boldsymbol{v}}_{t-1} + \begin{bmatrix}0 & 1\end{bmatrix} \begin{bmatrix} v_{t-1}^A(i) \\ v_{t-1}^{\alpha}(i) \end{bmatrix}$$

Making use of the previous-period quantity (Dixit-Stiglitz preferences)

Equivalent logic to the above then grants the following for Dixit-Stiglitz preferences:

$$\widehat{p}_{t-1}(i) + \widehat{y}_{t-1}(i) = (\boldsymbol{\gamma}_y + \varepsilon\boldsymbol{\gamma}_p)' X_{t-1} + \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} v^A_{t-1}(i) \\ v^\alpha_{t-1}(i) \end{bmatrix}$$

where ε is the elasticity of demand.

Including the wage in the signal vector

Recall from appendix 4.A.4 that

$$\widehat{\varpi}_t = \left(\frac{1}{\sigma} + \frac{1}{\psi}(1+\eta)\right)\widehat{y}_t - \frac{1}{\psi}(1+\eta)\widehat{z}_t$$
$$\widehat{z}_t = \epsilon^A_t + \overline{\widehat{p}_t} - \widehat{\mathbb{p}}_t$$

where $\widehat{\varpi}_t \equiv \widehat{w}_t - \widehat{\mathbb{p}}_t$ is the real wage. Combining these, we have

$$\widehat{w}_t = \boldsymbol{\gamma}'_w X_t + \boldsymbol{\delta}'_w \widetilde{\boldsymbol{v}}_t$$
$$\boldsymbol{\gamma}_w = \boldsymbol{\gamma}_p + \left(\frac{1}{\sigma} + \frac{1}{\psi}(1+\eta)\right)\boldsymbol{\gamma}_y - \frac{1}{\psi}(1+\eta)(\boldsymbol{\gamma}_{\overline{p}} - \boldsymbol{\gamma}_p)$$
$$\boldsymbol{\delta}_w = \left(\frac{1}{\sigma} + \frac{1}{\psi}(1+\eta)\right)\boldsymbol{\delta}_y + \left(1 - \frac{1}{\psi}(1+\eta)\right)\boldsymbol{\delta}_p$$

The signal vector

We are therefore able to write

$$\boldsymbol{q}_t(i) = C_1 X_t + C_2 X_{t-1} + R_1 \begin{bmatrix} v^A_t(i) \\ v^\alpha_t(i) \end{bmatrix} + R_2 \begin{bmatrix} v^A_{t-1}(i) \\ v^\alpha_{t-1}(i) \end{bmatrix} + R_3 \boldsymbol{e}_t + R_4 \widetilde{\boldsymbol{v}}_{t-1}$$

where

$$C_1 = \begin{bmatrix} & 0_{1\times\infty} & \\ & 0_{1\times\infty} & \\ & 0_{1\times\infty} & \\ & 0_{1\times\infty} & \\ 1 \quad 0 \quad 0 \quad 0 & 0_{1\times\infty} \end{bmatrix} \quad C_2 = \begin{bmatrix} \boldsymbol{\gamma}'_y \\ \boldsymbol{\gamma}'_p \\ \boldsymbol{\gamma}'_y + \boldsymbol{\gamma}'_p + \gamma\boldsymbol{\gamma}'_{\overline{p}} \\ \boldsymbol{\gamma}'_w \\ 0_{1\times\infty} \end{bmatrix} \quad R_1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}$$

$$R_2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad R_3 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad R_4 = \begin{bmatrix} \boldsymbol{\delta}'_y \\ \boldsymbol{\delta}'_p \\ \boldsymbol{\delta}'_y + \boldsymbol{\delta}'_p \\ \boldsymbol{\delta}'_w \\ 0 \quad 0 \end{bmatrix}$$

Innovation in the signal vector

Recall that the optimal linear estimator in this setting is the Kalman filter (4.68):

$$E_t(i)[X_t] = FE_{t-1}(i)[X_{t-1}] + K\{\boldsymbol{q}_t(i) - E_{t-1}(i)[\boldsymbol{q}_t(i)]\}$$

First note that firm i's prior expectation of $\boldsymbol{q}_t(i)$ (i.e. that formed in the previous period) may be written as

$$E_{t-1}(i)[\boldsymbol{q}_t(i)] = (C_1F + C_2)E_{t-1}(i)[X_{t-1}] + R_2E_{t-1}(i)\begin{bmatrix} v^A_{t-1}(i) \\ v^\alpha_{t-1}(i) \end{bmatrix}$$

where we have dropped the terms in $\boldsymbol{u}_t$, $v^A_t(i)$ and $v^\alpha_t(i)$ as they cannot be seen in advance and have unconditional expectations of zero. The term in $\widetilde{\boldsymbol{v}}_{t-1}$ is also dropped as idiosyncratic shocks are fully independent (so that $E_t(i)[\boldsymbol{v}_t(j)] = 0\ \forall j \neq i$) and in the limit, $\lim_{J\to\infty}\alpha(i)^* = 0\ \forall i$. It is also the case that $E_{t-1}(i)\left[v^\alpha_{t-1}(i)\right] = 0$ since firms make no observation of household demand at the moment of setting prices. In principle, we have a complication in that since the firm observes $\widehat{a}_t(i)$, it must be that

$$E_{t-1}(i)\left[v^A_{t-1}(i)\right] = v^A_{t-1}(i) + \epsilon^A_{t-1} - E_{t-1}(i)\left[\epsilon^A_{t-1}\right]$$

However, since nothing in $\boldsymbol{q}_t(i)$ depends on $E_{t-1}(i)\left[v^A_{t-1}(i)\right]$ (the left-hand column of R_2 is all zeros), we may ignore this complication and simply write

$$E_{t-1}(i)[\boldsymbol{q}_t(i)] = (C_1F + C_2)E_{t-1}(i)[X_{t-1}]$$

Consequently, we may write the surprise in (i.e. the unexplained portion of) the firm's signal vector as

$$\overrightarrow{\boldsymbol{q}}_t(i) \equiv \boldsymbol{q}_t(i) - E_{t-1}(i)\left[\boldsymbol{q}_t(i)\right]$$
$$= (C_1F + C_2)\left(X_{t-1} - E_{t-1}(i)\left[X_{t-1}\right]\right) + C_1G_1\boldsymbol{u}_t + R_1\begin{bmatrix} v_t^A(i) \\ v_t^\alpha(i) \end{bmatrix} + R_2\begin{bmatrix} v_{t-1}^A(i) \\ v_{t-1}^\alpha(i) \end{bmatrix}$$
$$+ (R_3 + C_1G_3)\,\boldsymbol{e}_t + (R_4 + C_1G_4)\,\widetilde{\boldsymbol{v}}_{t-1}$$

Bibliography

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.

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Saleh (2010): "Cascades in Networks and Aggregate Volatility," *NBER Working Papers*, No. 16516.

Adam, K. (2007): "Optimal monetary policy with imperfect common knowledge," *Journal of Monetary Economics*, 54(2), 267–301.

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

Alvarez, L., and I. Hernando (2005): "The Price Setting Behaviour of Spanish Firms: Evidence from Survey Data," *ECB Working Paper*, No. 538.

Amirault, D., C. Kwan, and G. Wilkinson (2006): "Survey of Price-Setting Behaviour of Canadian Companies," *Bank of Canada Working Paper*, No. 2006-35.

Apel, M., R. Friberg, and K. Hallsten (2005): "Microfoundations of Macroeconomic Price Adjustment: Survey Evidence from Swedish Firms," *Journal of Money, Credit, and Banking*, 37(2), 313–338.

Aucremanne, L., and M. Druant (2005): "Price-setting Behaviour in Belgium: What can be Learned from an Ad Hoc Survey?," *ECB Working Paper*, No. 448.

Axtell, R. L. (2001): "Zipf Distribution of U.S. Firm Sizes," *Science*, 293(September), 1818–1820.

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.

Bergin, P. R., and R. C. Feenstra (2000): "Staggered price setting, translog preferences, and endogenous persistence," *Journal of Monetary Economics*, 45(3), 657–680.

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

Bils, M., and P. Klenow (2004): "Some Evidence on the Importance of Sticky Prices," *Journal of Political Economy*, 112(5), 947–985.

Blackwell, D. (1965): "Discounted Dynamic Programming," *The Annals of Mathematical Statistics*, 36(1), 226–235.

Blanchard, O., and C. Kahn (1980): "The Solution of Linear Difference Models under Rational Expectations," *Econometrica*, 48(5), 1305–1311.

Blinder, A. (1991): "Why are Prices Sticky? Preliminary Results from an Interview Survey," *American Economic Review*, 81(2), 89–96.

Blinder, A., E. Canetti, D. Lebow, and J. Rudd (1998): *Asking about prices: a new approach to understanding price stickiness*. Russell Sage Foundation, New York, 1 edn.

Cai, G.-H. (2006): "Strong laws for weighted sums of i.i.d. random variables," *Electronic Research Announcements of the American Mathematical Society*, 12, 29–36.

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.

CARVALHO, V. (2010): "Aggregate Fluctuations and the Network Structure of Intersectoral Trade," mimeo.

CHARI, V., P. KEHOE, AND E. MCGRATTAN (2000): "Sticky price models of the business cycle: Can the contract multiplier solve the persistence problem?," *Econometrica*, 68(5), 1151–1180.

CHRISTENSEN, L., D. JORGENSON, AND L. TAU (1975): "Transcendental Logarithmic Utility Functions," *American Economic Review*, 65(3), 367–383.

CHRISTIANO, L., M. EICHENBAUM, AND C. EVANS (1999): "Monetary Policy Shocks: What Have We Learned and to What End?," in *Handbook of Macroeconomics*, ed. by J. Taylor, and M. Woodford, vol. 1A, chap. 2, pp. 65–148. Elsevier, New York.

CLAUSET, A., C. R. SHALIZI, AND M. E. J. NEWMAN (2009): "Power-Law Distributions in Empirical Data," *SIAM Review*, 51(4), 661–703.

CORLESS, R. M., AND D. J. JEFFREY (2002): "The Wright ω function," in *Artificial Intelligence, Automated Reasoning, and Symbolic Computation*, ed. by J. Calmet, B. Benhamou, O. Caprotti, L. Henocque, and V. Sorge, pp. 76–89. Springer.

DEATON, A., AND J. MUELLBAUER (1980): "An Almost Ideal Demand System," *American Economic Review*, 70(3), 312–326.

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.

DIXIT, A., AND J. STIGLITZ (1977): "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67(3), 297–308.

DOSSCHE, M., F. HEYLEN, AND D. VAN DEN POEL (2010): "The Kinked Demand Curve and Price Rigidity: Evidence from Scanner Data," *Scandinavian Journal of Economics*, 112(4), 723–752.

Fabiani, S., M. Druant, I. Hernando, C. Kwapil, B. Landau, C. Loupias, F. Martins, T. Mathä, R. Sabbatini, H. Stahl, and A. Stokman (2005): "The Pricing Behaviour of Firms in the Euro Area: New Survey Evidence," *ECB Working Paper*, No. 535.

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.

Gertler, M., and J. Leahy (2008): "A Phillips Curve with an Ss Foundation," *Journal of Political Economy*, 116(3), 533–572.

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

Grewel, M. S., and A. P. Andrews (2008): *Kalman Filtering: Theory and Practice Using MATLAB*. John Wiley & Sons, Inc., Hoboken, New Jersey, 3 edn.

Hall, S., M. Walsh, and A. Yates (1997): "How do UK companies set prices?," *Bank of England Working Paper*, No. 67.

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

Hoeberichts, M., and A. Stokman (2010): "Price Setting Behaviour in the Netherlands: Results of a Survey," *Managerial and Decision Economics*, 31(2), 135–149.

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

Kennan, J. (2001): "Uniqueness of Positive Fixed Points for Increasing Concave Functions on R^n: An Elementary Result," *Review of Economic Dynamics*, 4(4), 893–899.

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

KIMBALL, M. (1995): "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit, and Banking*, 27(4), 1241–1277.

KLEIN, P. (2000): "Using the generalized Schur form to solve a multivariate linear rational expectations model," *Journal of Economic Dynamics and Control*, 24(10), 1405–1423.

KLENOW, P., AND O. KRYVTSOV (2008): "State-Depenent or Time-Dependent Pricing: Does It Matter For Recent U.S. Inflation?," *Quarterly Journal of Economics*, 123(3), 863–904.

KLENOW, P., AND B. MALIN (2010): "Microeconomic Evidence on Price-Setting," in *Handbook of Monetary Economics*, ed. by B. Friedman, and M. Woodford, vol. 3, chap. 6, pp. 231–284. Elsevier.

KWAPIL, C., J. BAUMGARTNER, AND J. SCHARLER (2005): "The Price-Setting Behavior of Austrian Firms: Some Survey Evidence," *ECB Working Paper*, No. 464.

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.

LÜNNEMANN, P., AND T. MATHÄ (2006): "New Survey Evidence on the Pricing Behaviour of Luxembourg Firms," *ECB Working Paper*, No. 617.

LOUPIAS, C., AND R. RICART (2004): "Price Setting in France: New Evidence from Survey Data," *ECB Working Paper*, No. 423.

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

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.

Martins, F. (2005): "The Price Setting Behaviour of Portuguese Firms: Evidence from Survey Data," *ECB Working Paper*, No. 562.

Melosi, L. (2012): "Signaling Effects of Monetary Policy," *Federal Reserve Bank of Chicago Working Paper*, No. 2012-05.

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

Mortensen, D., and C. Pissarides (1994): "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies*, 61(3), 397–415.

Muellbauer, J. (1975): "Aggregation, Income Distribution and Consumer Demand," *Review of Economic Studies*, 42(4), 525–543.

——— (1976): "Community Preferences and the Representative Consumer," *Econometrica*, 44(5), 979–999.

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

Nakagawa, S., R. Hattori, and I. Takagawa (2000): "Price-Setting Behavior of Japanese Companies," *Bank of Japan Research Papers*, No. 2000-09.

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.

Peersman, G., and F. Smets (2003): "The Monetary Transmission Mechanism in the Euro Area: More Evidence from VAR Analysis," in *Monetary Policy Transmission in the Euro Area*, ed. by I. Angeloni, A. Kashyap, and B. Mojon, chap. 2, pp. 36–55. Cambridge University Press.

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.

Plosser, C. (2012): "Macro Models and Monetary Policy Analysis," Speech given at Bundesbank – Federal Reserve Bank of Philadelphia Spring 2012 Research Conference.

Reis, R. (2006): "Inattentive Producers," *Review of Economic Studies*, 73(3), 793–821.

Romer, D., and C. Romer (2004): "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94(4), 1055–1084.

Rotemberg, J. (1982): "Sticky Prices in the United States," *Journal of Political Economy*, 90(6), 1187–1211.

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

Singleton, K. (1987): "Asset prices in a time series model with disparately informed, competitive traders," in *New Approaches to Monetary Economics*, ed. by W. Burnett, and K. Singleton, vol. 1. Cambridge University Press.

Smets, F., and R. Wouters (2003): "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1(5), 1123–1175.

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

STAHL, H. (2005): "Price Setting in German Manufacturing: New Evidence from New Survey Data," *ECB Working Paper*, No. 561.

SUNG, S. H. (2001): "Strong laws for weighted sums of i.i.d. random variables," *Statistics and Probability Letters*, 52(4), 413–419.

TAYLOR, J. (1980): "Aggregated Dynamics and Staggered Contracts," *Journal of Political Economy*, 88(1), 1–23.

——— (1999): "Staggered price and wage setting in macroeconomics," in *Handbook of Macroeconomics*, ed. by J. Taylor, and M. Woodford, vol. 1, pp. 1009–1050. Elsevier, 1 edn.

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.

WU, W. B. (1999): "On the strong convergence of a weighted sum," *Statistics and Probability Letters*, 44(1), 19–22.