Essays on Expectation Formation

Katharine Lauderdale

University College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Michaelmas 2017

Abstract

This thesis is composed of three independent papers that model patterns of heterogeneity in inflation expectations and price-setting.

In the first paper, I present stylized facts from the Survey of Professional Forecasters that suggest the inflation expectation formation process changes over the term of the forecast. I develop a model to capture these features: forecasters have (1) a cost for inacccuracy, for which they pay the penalty when the variable is realized and (2) a cost for inconsistency (changing a forecast from quarter to quarter), for which they pay the penalty when the adjustment is reported. Due to heterogeneous time discounting, the inconsistency cost is more highly but differentially weighted at long horizons, which induces the observed stylized facts.

In the second paper, I estimate a sector-level Rational Inattention (RI) price-setting model. I estimate the model separately on two regimes (pre-Volcker and Great Moderation) to investigate predictions from RI theory about price-setting reactions to a change in volatility. I find that the model is not capable of reproducing price level dynamics, and that the attention capacity has low predictive power across regimes. Given these caveats, there is some evidence that attention capacity increased across the sample.

In the third paper, I consider two kinds of uncertainty: the mean and the variance of the prior distribution of a stochastic variance. I build a model with strategic complementarity and provide comparative statics welfare results. Higher mean-uncertainty in the private signal always decreases welfare, and higher mean-uncertainty in the public signal sometimes decreases welfare. However, higher variance-uncertainty in the private signal always increases welfare, and higher variance-uncertainty in the public signal sometimes increases welfare. This finding contributes to the debate as to whether this class of model supports limited transparency among policy makers.

Essays on Expectation Formation



Katharine Lauderdale
University College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Michaelmas 2017

Acknowledgements

I would like to thank Sophocles Mavroeidis for his excellent guidance, support, advice, and patience.

I am grateful for comments from Martin Ellison, Franck Portier, John Muellbauer, Stephen Hansen, Guillaume Chevillon, and the audience of my Gorman workshop. Particular thanks go to Ben Deaner for reading and critiquing pieces of this thesis at various points in its development and to Fergus Cooper for technical assistance.

Lastly, warm thanks to Diane, two Bens, and Maeve for their respective varieties of optimism.

Contents

List of Figures **v**

List of Tables **vi**

1 Introduction **1**

2 Disagreement and consistency among professional forecasters **8**

2.1 Introduction 8
2.2 Background 10
2.3 Data 12
2.4 Model 13
2.4.1 State space 17
2.4.2 Illustration of the model 19
2.5 Results 20
2.6 Conclusion 22

3 Estimation of a Rational Inattention model with multiple volatility regimes **28**

3.1 Introduction 28
3.2 Information theory 31
3.3 Motivating model 32
3.4 Optimal signal choice 35
3.4.1 Univariate optimal signal 35
3.4.2 Bivariate optimal signal 41
3.5 Volatility regimes 44
3.6 Results 45
3.6.1 Data 45
3.6.2 Estimation 46
3.6.3 Regime definition 47
3.6.4 ARMA estimates 48
3.6.5 Attention estimates 51
3.7 Conclusion 58

4 Social value of uncertainty **61**
4.1 Introduction . . . 61
4.2 Model . . . 64
4.2.1 MS model . . . 64
4.2.2 Generalized model . . . 67
4.3 Welfare results . . . 70
4.4 Example . . . 74
4.5 Conclusion . . . 79

5 Conclusion **81**

Appendices

A Appendix: Disagreement and consistency among professional forecasters **85**
A.1 Stylized facts for the noisy information model . . . 85
A.2 Stylized facts for the consistency objective . . . 86
A.3 GMM moments . . . 88
A.3.1 Inflation process moments . . . 88
A.3.2 Consistency parameter moments . . . 88
A.3.3 Cross-sectional variance moments . . . 89

B Appendix: Estimation of a Rational Inattention model with multiple volatility regimes **91**
B.1 Proof of Proposition 3.1 . . . 91
B.2 Proof of Proposition 3.2 . . . 92
B.3 Proof of Proposition 3.3 . . . 93
B.4 Proof of Proposition 3.4 . . . 93
B.5 Proof of Proposition 3.5 . . . 95
B.6 Proof of Proposition 3.6 . . . 96
B.7 Proof of Proposition 3.7 . . . 99
B.8 Proof of Proposition 3.8 . . . 100
B.9 Personal Consumption Expenditure goods indices . . . 102

C Appendix: Social value of uncertainty **104**
C.1 Model definitions . . . 104
C.2 Equilibrium existence . . . 105
C.3 Model properties . . . 108
C.3.1 Differentiability assumptions . . . 108
C.3.2 Model properties . . . 108
C.4 Welfare results . . . 111

Bibliography **118**

List of Figures

2.1 The optimal reporting weight by horizon and discount rate 23
2.2 Simulated inflation process and forecasts 24
2.3 Discount rate estimates 25
2.4 Regression of disagreement on inflation level 26
2.5 Regression of disagreement on inflation change 27

3.1 Disaggregated Personal Consumption Expenditure indices. 47
3.2 Mean ex post variance reduction for different regime definitions . . 49
3.3 Aggregate series in each regime 50
3.4 Impulse responses to a one a standard deviation nominal demand shock 52
3.5 Ex post variance reduction by regime 54
3.6 Change in ex post variance reduction between regimes 55
3.7 Regime 1 ex post variance reduction with labeled PCE sectors . . . 56
3.8 Change between regimes in ex post variance reduction with labeled PCE sectors 57
3.9 Correlation in ex post variance reduction between regimes 59

4.1 The prior distribution of the stochastic variance for different levels of mean- and variance-uncertainty 70
4.2 The distribution of the shock with varying mean- and variance-uncertainty 74
4.3 The agent's conditional expectation with varying mean- and variance-uncertainty 75
4.4 The action schedule for varying mean- and variance-uncertainty (low beauty-contest weight) 76
4.5 The action schedule for varying mean- and variance-uncertainty (high beauty-contest weight) 77
4.6 The expected welfare for varying mean- and variance-uncertainty (low beauty-contest weight) 78
4.7 The expected welfare for varying mean- and variance-uncertainty (high beauty-contest weight) 79

List of Tables

2.1 Disagreement regressed on change in and level of inflation 14
2.2 Disagreement inheritance between periods 15
2.3 GMM estimates of model parameters 21
2.4 Bias and SD of GMM estimates 21

3.1 ARMA estimates 51
3.2 Mean ex post variance reduction estimates in each regime 53
3.3 Variance decomposition by regime 53

4.1 The effect of mean- and variance-uncertainty on welfare terms . . . 71

1 Introduction

In recent years, there has been an increased regard among macroeconomists for the key roles of information and expectation formation in macroeconomic dynamics. This is reflected in both the policy choices of governments since the 2008 crisis, as well as in new developments in academic economics. Expectations about the future play a role in a wide variety of decisions, such as consumption, savings, price-setting, wage-setting, and hiring. Policies of forward guidance are built on the idea that expectations play an important role in shaping the path of real and nominal variables. Fiscal stimulus packages are justified in modern Keynesian models with depressed aggregate demand as a function of agents' forward expectations. In particular, inflation expectations are important in models with a Phillips curve relationship between unemployment and inflation (Mankiw and Reis, 2001), or models in which inflation expectations affect aggregate demand through consumption and investment, in the vein of Clarida et al. (1999).

Imperfect information models can be used to capture a variety of stylized facts, but a key feature is dampened and delayed response functions of prices and quantities. It is particularly important to understand this inertia in the case of price adjustment, in order to understand the mechanisms of real short-run effects of monetary policy. Keynesian models posit price stickiness as the lever of monetary policy, and adjustment costs for quantity variables can be added to model sluggish

production responses. Models of imperfect information can also achieve these delayed price and quantity responses (Adam 2007, Paciello and Wiederholt 2013, Roca 2010). These models can induce serially correlated responses even if the underlying process being tracked has little persistence. In models where agents have strategic interactions, this effect can be amplified, as agents react not only to noise but also to expectations of others' expectation (Angeletos and La'O 2013, Amato and Shin 2003). Imperfect information theories can explain multiple patterns of inertial responses under a unified theory, such as explaining sector-level price volatility but aggregate inertia (Boivin et al. 2009, Maćkowiak and Wiederholt 2009, Maćkowiak et al. 2009), or explaining different sensitivities of output to changes in supply and demand (Maćkowiak and Wiederholt, 2015). Lastly, the idea of imperfect information intuitively accords with our individual experience as economic agents with limited macroeconomic knowledge. It seems a priori plausible that the effects of this individual ignorance aggregate to observable effects.

There have been a range of approaches to modeling expectation formation processes, and they make varying assumptions about agents' informedness and computational sophistication. At one extreme are simple models of adaptive expectations, and at the other extreme are full information rational expectations models. The models in this thesis place constraints on agent's information sets. Conditional on these limitations, agents have no constraints on their computational capacity and choose their loss-minimizing response.

In the rest of the introduction, I discuss relevant background for each of the three substantive chapters of this thesis. Although the three chapters are independent and can be read separately, all are concerned with explaining patterns of heterogeneity in inflation expectations and price-setting. Macroeconomic models are often constructed with homogenous firms and consumers. As the dynamics of an aggregate series can often be captured using many different information structures, the cross-sectional dispersion can provide additional evidence about departures from full-information rational expectations behavior (Coibion and Gorodnichenko, 2012). Moreover, dispersion in agent information has implications for the efficacy

of particular policies that seek to anchor expectations. The efficacy of inflation targeting for reducing expectation dispersion has been the focus of much of this literature (Crowe 2010, Eijffinger and Geraats 2006).

Chapter 2: Disagreement and consistency among professional forecasters

Survey data of macro forecasts can be used to investigate a range of questions about expectations formation processes. The data can test predictive accuracy of forecasts, estimate the underlying loss function of agents and the information content of private signals (Mankiw et al. 2003). The term structure of disagreement and correlation of forecast errors provide evidence about agents' filtering mechanism for distinguishing between persistent and transitory shocks and noise (Coibion and Gorodnichenko 2015, Patton and Timmermann 2010) and also reflects the kind of structural model agents use for multidimensional updating across multiple series (Atkeson and Ohanian 2001, Dovern 2014). There are theories competing with standard Gaussian signal-extraction models about the causes of disagreement: sticky information (Mankiw and Reis 2001, Mankiw et al. 2003), heterogeneous priors (Andrade et al. 2013, Ashiya 2009), and heterogeneous loss functions (Elliott et al. 2008, Capistrán and Timmermann 2009, Ehrbeck and Waldmann 1996, Patton and Timmermann 2007). Agents may pay attention to news sources that degrade the quality of informtion (Carroll, 2003). In particular, it is of interest to characterize how the deviations from survey data from rational expectations theory affect aggregate expectations and policy efficacy. Time-variation in disagreement provides evidence about state-dependent expectation dispersion for evaluating policies at different points in the state space (Dovern et al., 2012).

Chapter 2 seeks to explain patterns of disagreement in forecaster survey data. First, I present new stylized facts from the Survey of Professional Forecasters that suggest the expectation formation process for inflation changes over the term of the forecast. Change in inflation and the level of inflation correspond to larger increases in disagreement at longer forecast horizons, and disagreement is more

highly predictive of disagreement in the next period at longer horizons. I suggest a model that captures these features, by augmenting forecasters' loss function. In the model, forecasters have (1) a cost for inacccuracy, for which they pay the penalty when the true value of the variable is realized and (2) a cost for inconsistency (changing a forecast from quarter to quarter), for which they pay the penalty when the adjustment is reported. Due to heterogeneous time discounting, the inconsistency cost is more highly but differentially weighted at long horizons, which induces the observed stylized facts.

Chapter 3: Estimation of a Rational Inattention model with volatility regimes

Chapter 3 estimates a model from the Rational Inattention (RI) literature, a theory originally articulated in Sims (1998) and Sims (2003). In this framework, agents are subject to constraints on information flow, measured as Shannon entropy. This theory provides an elegant explanation for nominal rigidity and gradual adjustment of macroeconomic variables, despite the great quantity of economic information available (Steiner et al., 2015). RI provides a defense for demand-driven models of the business cycle in the face of evidence that menu costs are small. If one thinks of sticky prices not as literal menu costs but rather as an approximation of some computational cost to changing prices, then RI models provide a congruent alternative. RI models have additional attractive features. Information constraints do not create the sticky-price modeling phenomenon of anticipatory price changes in advance of a rise in nominal output (Woodford 2003). RI models suggest that high volatility can lower welfare because agents must make decisions with worse information than they would under low volatility. Information constraints may be different for different kinds of information, and so agents may be more or less reactive to different kinds of signals. By contrast, Calvo-type pricing induces one constraint for price responses to all kinds of shocks. For example, RI can explain prices being more responsive to changes in factor prices than changes in demand, as agents may have more information about supply cost (Maćkowiak and Wiederholt,

2015). Lastly, RI accords with results from behavioral work, experimental and survey data, that show numerous forms of cognitive limitations (Sims et al. 2005).

The RI literature has been largely theoretical, and Chapter 3 seeks to estimate a RI model (Maćkowiak and Wiederholt 2009) on price data. RI theory makes predictions about how agents will reallocate their attention under different volatility regimes. I investigate whether these patterns are detectible in the data by estimating the model separately in a high-volatility pre-Volcker regime and a low-volatility Great Moderation regime (regimes documented in, for example, Kim and Nelson 1999, McConnell and Perez-Quiros 2000, and Stock and Watson 2003). The goal is to determine whether attention capacity in one regime is related to attention capacity in the second regime. Although the estimation produces a similar distribution of attention parameters in both volatility regimes, the attention parameters are weakly identified and have large confidence intervals. The estimation is subject to the caveat that the model is unable to reproduce the aggregate price level dynamics in the data, perhaps due to the presence of only one aggregate shock. Given these limitations, the estimation suggests that firms have on average increased capacity for information processing over the span of the sample.

Chapter 4: Social value of uncertainty

In the imperfect information literature, it is often unclear when it is reasonable to assume that agents know the true data-generating process. Some models have agents with heterogeneous priors, or the effect of their baseline priors does not dissipate with time (Patton and Timmermann, 2010). One of the most important components of the prior is the volatility, which is commonly assumed to be known by the agents. Formally relaxing this assumption is a logical extension of the research agenda of imperfect information.

This issue is particularly relevant for the literature on uncertainty shocks. The idea that changes in the level of volatility cause business cycle fluctuations has intuitive and anecdotal appeal. In an analogous way to news shocks (Beaudry and Portier 2014), uncertainty shocks could account for movements that are difficult to

explain with measurable changes to fundamentals (Berger et al., 2017). There has been empirical evidence and theoretical work on uncertainty about productivity (Bloom 2009, Bloom et al. 2012, Fernández-Villaverde et al. 2011, Nimark 2014). There is also a literature finding effects of fiscal policy uncertainty (Born and Pfeifer 2014, Fernández-Villaverde et al. 2015, Pastor and Veronesi 2012, Pástor and Veronesi 2013), particularly in sectors with large shares of government spending (Baker et al. 2016). Policy uncertainty may induce wait-and-see-type behavior, given political disagreement over fiscal spending and monetary policy controversies over quantitative easing programs. But other studies have found insignificant effects of uncertainty shocks (Basu and Bundick 2017, Bachmann et al. 2013, Bachmann and Bayer 2011, Bekaert et al. 2013, Popescu and Rafael Smets 2010). These studies have found uncertainty mechanisms working in cancelling directions. Bloom et al. (2012) found that uncertainty shocks initially make policies initially less effective and then subsequently more effective. One possible reason for contradictory findings is that agents have uncertainty about the level of uncertainty; not only do they have uncertainty with respect to future volatility, but they also do not know the current volatility. It is possible that investigation into the effects of variance-uncertainty would shed light on the disagreements in the literature about the effects of uncertainty shocks.

I investigate a basic theoretical aspect of this issue: how variance-uncertainty behaves in a model with strategic complementarity. I use the widely-cited model in Morris and Shin (2002). Their arguments formalize the concern among policymakers that public signals can cause socially suboptimal herding (for a review, see Blinder et al. 2008). If agents have decentralized information, then they will have expectations of other agents' expectations. As a result, the public signal plays a dual role of providing precise information and also guiding the agent's strategy. This class of model, in which agents with heterogeneous information sets have coordination failures in price-setting around noisy public signals, can enhance the patterns of delay and persistance to monetary policy reactions (Angeletos and

Pavan 2004, Angeletos and Pavan 2007). Higher-order expectations can display even more inertia than simple linear filtering reaction functions.

In Chapter 4, I compare the effects of two kinds of uncertainty shocks. The first is the typical definition from the literature, the variance of a shock series. This is uncertainty with respect to the mean of a shock, which I compare to uncertainty with respect to the variance of a shock. Higher variance-uncertainty captures the idea that extreme (low or high) levels of volatility become more probable, holding expected volatility constant. These two kinds of uncertainty can be parameterized respectively as the mean and the variance of the prior distribution of the stochastic variance. I provide comparative statics results showing that the two kinds of uncertainty have contrasting welfare implications in this model. Higher mean-uncertainty in the private signal always decreases welfare, and higher mean-uncertainty in the public signal sometimes decreases welfare. However, higher variance-uncertainty in the private signal always increases welfare, and higher variance-uncertainty in the public signal sometimes increases welfare.

2

Disagreement and consistency among professional forecasters

2.1 Introduction

The press coverage of macroeconomic forecasting frequently discusses the cost of admitting to modeling errors.[1] This journalistic angle implies that the forecaster incurs a loss not simply for inaccurate estimates, but also for publicly changing an opinion in light of new information. This chapter aims to model this idea in order to capture features of survey forecaster data.

A central feature of inflation forecasting is that disagreement rises with both the level and the change in inflation, and these patterns are well-documented in the literature (Dovern et al. 2012, Mankiw et al. 2003). To model these features, I present a model of expectations formation based on a noisy information model but augmented by a simple model of forecaster behavior. Forecasters have a two-part objective. First, they face quadratic loss for innacurate forecasts. This

[1]Recent examples include, "Inflation Unfriends the Fed" (*Wall Street Journal* 13/10/2017); "Central Bankers have one job and they don't know how to do it" (*Financial Times* 18/10/2017); a solid majority of Paul Krugman columns, most recently "The Doctrine of Trumpal Infallibility" (*New York Times* 23/10/2017). More general discussions include "To err is human; so is the failure to admit it" (*Economist* 10/6/2017); "Brexit experts, confess to your errors and carry on" (*Financial Times* 4/1/2017); "Fed Critics Say '10 Letter Warning Inflation Still Right" (*Bloomberg* 2/10/2014).

cost is incurred in the future, when inflation is realized and the error can be verified. Second, forecasters face a quadratic loss when they adjust their forecast of a particular event from quarter to quarter: an *inconsistency* cost. This cost is incurred in the period when the adjustment to the forecast is made. The consistency objective could be intuitively explained in several ways: there may be adjustment costs for the forecaster's firm switching strategies, or there may be a reputational cost to the forecaster for admitting to previous inaccuracy. Forecasters have heterogeneous time discount rates, meaning that the forecasters with relatively steep discounting weigh the immediate cost of inconsistency more heavily and the distant cost of innacuracy more lightly.

The model contains more specific predictions. Due to the timings of the inaccuracy and inconsistency costs, they have different relative effects depending on the horizon length. At short horizons, the cost of inaccuracy will be incurred relatively soon, and so forecasters prioritize maximizing accuracy rather than maximizing consistency. At long horizons, the cost of inaccuracy will be more highly time-discounted, so maximizing consistency is weighted relatively highly.

Because these weights will vary depending on the forecasters' heterogeneous discount rates, there are implications for patterns of disagreement. Heterogeneity in discount rates is effectively exponentially magnified as the horizon increases. Consequently, at short horizons, there will be relatively low heterogeneity and high weighting of the accuracy objective. Forecasters' objective functions will be close to a simple quadratic loss for inaccuracy, and so their behavior will be more similar to a noisy information model. Hence, the effect of the level and change in inflation will be relatively small, and disagreement will be weakly inherited from the previous period. At long horizons, forecasters will weigh consistency relatively highly, and their objective functions will become more heterogeneous. A change in inflation will then be differentially incorporated into forecasters' updates (some will respond slowly to new information), increasing disagreement. Similarly, a level of inflation that deviates from the long run mean of the inflation process will be heterogeneously

incorporated into forecasts, also increasing disagreement. Additionally, at long horizons, disagreement will be relatively strongly inherited from the previous period.

Using the Survey of Professional Forecasters (SPF), I present stylized facts that show that these patterns of disagreement are indeed in the data. While both the level and the change in inflation significantly predict disagreement at every horizon length, they have larger effects at longer horizons. Additionally, lagged disagreement more strongly explains variation in disagreement at longer horizons. I fit the model to the SPF data via GMM and provide fit diagnostics.

This chapter is motivated in two ways. First, it investigates the prevalent idea that there is a cost associated with producing inconsistent macroeconomic forecasts. Models of forecasting have not previously used a consistency objective, nor heterogenous discount rates, nor have they had objective functions that depend on the horizon length.[2] Second, there is an empirical motivation, which is to investigate the drivers of forecaster disagreement given the great volume of public macroeconomic data. In particular, it is focused on explaining new stylized facts about the term structure of disagreement.[3]

In Section 2.2, I summarize recent literature on information rigidities and forecaster behavior. In Section 2.3, I provide the stylized facts from the SPF. In Section 2.4, I present the model specification and use simulated data to illustrate the mechanism of the model. In Section 2.5, I present empirical results. Section 2.6 concludes.

2.2 Background

Inflation is difficult to forecast, and the nature of the forecasting problem has changed over time. Since the early 1980s, inflation has become much less volatile, reducing the magnitude of forecasting errors. However, during this period of

[2]Previous papers with behavioral models of forecasters have instead focussed on herding (see Coibion and Gorodnichenko 2012, Pons-Novell 2003, and Pierdzioch and Rülke 2012) and differentiation (Ashiya 2009 and Tillmann 2011).

[3]The only previous paper, so far as I know, that uses the term structure of disagreement is Patton and Timmermann (2010), which simply notes that disagreement is higher at longer horizons.

moderation, multivariate forecasting models using the Phillips Curve fail to do any better than naive, univariate random walk forecasts (Stock and Watson 2007 and Atkeson and Ohanian 2001). This fact leads to the empirical question of how firms and professional forecasters deviate from naive forecasts, and what kinds of information rigidity are salient for the inflation expectation formation process.

The most prominent classes of models on information rigidity are the noisy information signal-extraction model (on which Rational Inattention models (Sims, 2003) are typically based) and the sticky information model (Mankiw and Reis, 2001). In the noisy information model, agents each receive a noisy signal of a macroeconomic variable. They then filter their information to form an estimate of the true state of the economy. In the sticky information model, agents update their information set with a fixed probability in every period, but when they do update their information set it is uncorrupted by noise. In addition, other models of forecaster behavior are the heterogeneous priors model (Patton and Timmermann, 2010) and the heterogeneous asymmetric loss function model (Capistrán and Timmermann, 2009). In the heterogeneous priors model, forecasters receive a signal with common and individual noise, but they also have a fixed prior about the long-run mean of the variable, which is incorporated into their filtering of new information. In the heterogeneous asymmetric loss function model, individual forecasters do not face quadratic loss for mispredicting an event; they may face greater penalties for over- or under-prediction.

All of the above models predict similar behavior for the consensus (mean) forecast error. Because forecasters are filtering to distinguish between (1) noise, (2) persistent shocks, and (3) white noise shocks, they will not immediately adjust their forecasts to new information. This will induce serially correlated consensus forecast errors. Coibion and Gorodnichenko (2012) show that all of the above models predict this behavior, and argue that patterns of disagreement (cross-sectional variance) in forecasts should be used to distinguish the applicability of various models.

There is an additional well-documented stylized fact which is not captured by the models above. Inflation disagreement rises not only with the change in inflation

but also with the level of inflation (see Capistrán and Timmermann 2009, Carroll 2003, Mankiw et al. 2003, Cukierman and Wachtel 1979, Valev and Carlson 2003, and Souleles 2004). I document this pattern in the following section.

2.3 Data

The Survey of Professional Forecasters (SPF) is a quarterly survey collected by the Philadelphia Federal Reserve. The sample size of forecasters varies from nine to forty, and the forecasts are collected over a variety of horizons. I focus on inflation as measured by GDP deflator (GNP deflator before 1992), a variable which has been surveyed at all the relevent horizon lengths since 1981Q3. The SPF asks both for forecasts of the quarter-over-quarter and year-over-year rates. The quarterly forecasts are for horizons of 1 to 4 quarters, and the annual forecast is for the annual average of the next calendar year (Jan 1 - Dec 31). That is, in 2000Q1, forecasters reported estimates for 2000Q2, 2000Q3, 2000Q4, 2001Q1, and mean inflation for 2001 (average of 2001Q1 to 2001Q4). In the analysis below, the 1- to 4-quarter horizon forecasts are from the quarterly forecasts. The 5- to 7-quarter horizons refer to the annual forecasts, converted to the implied average quarterly rate, made 5 to 7 quarters ahead of the end of the next calendar year.[4]

I present stylized facts on three predictors of forecaster disagreement: (1) the deviation of inflation from its long-run mean, (2) the change in inflation, and (3) the disagreement in the previous quarter.

Table 2.1 documents that disagreement rises with the level and change in inflation. Let y_t refer to demeaned inflation, $|y_t|$ to the deviation from the mean, and $|\Delta y_t|$ to the magnitude of the change in inflation. Inflation was demeaned using the full sample mean, as we are interested in the long-run mean of the series. Absolute values are used for empirical and conceptual reasons. I am investigating how forecasters incorporate new information, which should enter the forecaster's information set

[4]Unfortunately, the SPF does not contain longer horizons, as they have only recently started collecting 5- and 10-year forecasts. However, Lahiri and Sheng (2008) report that 2-year and 10-year forecasts for inflation are very similar.

symmetrically with respect to the sign. Reassuringly, when when these regressions were run with signed variables, the effects were smaller than with absolute values.

Let $\widehat{y}_{t|t-h}$ refer to the forecast of inflation at time t reported at time $t-h$. Let $\widehat{\mathrm{SD}}(\widehat{y}_{t|t-1})$ to $\widehat{\mathrm{SD}}(\widehat{y}_{t|t-4})$ refer to the sample cross-sectional standard deviation (disagreement) of 1 to 4 quarter horizon forecasts, and let $\widehat{\mathrm{SD}}(\widehat{y}_{t|t-5})$ to $\widehat{\mathrm{SD}}(\widehat{y}_{t|t-7})$ refer to forecasts of annual rates made 5 to 7 quarters before the final quarter of the forecasted year.

Disagreement for each horizon, $\widehat{\mathrm{SD}}(\widehat{y}_{t|t-h})$, was regressed on an intercept, the change in inflation $|\Delta y_{t-h}|$ and the deviation from the mean $|y_{t-h}|$ in the period when the forecast was made. Newey-West standard errors were calculated for all analyses in this section. Note that the effect of the change in inflation is positive at all horizons and rises monotonically with the horizon length, from 0.26 for a one-quarter forecast to 2.24 for the longest annual forecast. The deviation from inflation from its mean exhibits a similar pattern; the effects are all positive and rise with the horizon, from 0.03 at the shortest horizon to 0.84 at the longest horizon.

Table 2.2 documents the inheritance of disagreement from period to period. Disagreement in period $t-h$ about inflation in period t was regressed on an intercept and disagreement in the previous period, $t-h-1$, about inflation in period t. The R^2 coefficients for these regressions change over the term. For the quarterly forecasts with short horizons, the R^2 is relatively low, around 0.35. For the annual forecasts with longer horizons, the R^2 is higher, between 0.6 and 0.75. This provides rough evidence that disagreement is more persistent at longer horizons.

2.4 Model

This section first presents the model specification and then informally illustrates the mechanisms that induce the stylized facts shown above.

Inflation follows a reduced-form model that is consistent with Patton and Timmermann (2010). Inflation, y_t, has a latent stationary AR(1) process x_t, with white noise, u_t:

LHS	RHS	Estimate	Std. Error	t value	p
$\widehat{\text{SD}}(\widehat{y}_{t\|t-1})$	(Intercept)	-0.026	0.016	-1.574	0.118
	$\|\Delta y_{t-1}\|$	0.262	0.079	3.310	0.001
	$\|y_{t-1}\|$	0.028	0.016	1.742	0.084
$\widehat{\text{SD}}(\widehat{y}_{t\|t-2})$	(Intercept)	-0.049	0.023	-2.110	0.037
	$\|\Delta y_{t-2}\|$	0.356	0.120	2.974	0.004
	$\|y_{t-2}\|$	0.051	0.023	2.277	0.024
$\widehat{\text{SD}}(\widehat{y}_{t\|t-3})$	(Intercept)	-0.098	0.047	-2.083	0.039
	$\|\Delta y_{t-3}\|$	0.572	0.194	2.943	0.004
	$\|y_{t-3}\|$	0.100	0.046	2.192	0.030
$\widehat{\text{SD}}(\widehat{y}_{t\|t-4})$	(Intercept)	-0.114	0.040	-2.863	0.005
	$\|\Delta y_{t-4}\|$	0.687	0.175	3.928	<0.001
	$\|y_{t-4}\|$	0.116	0.039	2.990	0.003
$\widehat{\text{SD}}(\widehat{y}_{t\|t-5})$	(Intercept)	-0.390	0.191	-2.038	0.051
	$\|\Delta y_{t-5}\|$	0.404	0.443	0.911	0.370
	$\|y_{t-5}\|$	0.386	0.188	2.055	0.049
$\widehat{\text{SD}}(\widehat{y}_{t\|t-6})$	(Intercept)	-0.685	0.207	-3.303	0.003
	$\|\Delta y_{t-6}\|$	2.133	0.755	2.827	0.008
	$\|y_{t-6}\|$	0.673	0.202	3.323	0.002
$\widehat{\text{SD}}(\widehat{y}_{t\|t-7})$	(Intercept)	-0.856	0.191	-4.473	<0.001
	$\|\Delta y_{t-7}\|$	2.245	0.678	3.310	0.003
	$\|y_{t-7}\|$	0.841	0.187	4.496	<0.001

Table 2.1: Regressions of disagreement on the unsigned change in inflation and the unsigned deviation of inflation from its mean. The specification is repeated for horizon lengths of one to seven quarters.

$$y_t = x_t + u_t \tag{2.1}$$

$$x_t = \phi x_{t-1} + \epsilon_t, \quad -1 < \phi < 1 \tag{2.2}$$

Shocks are distributed

$$u_t \sim iid(0, \sigma_u^2), \quad \forall t$$

$$\epsilon_t \sim iid(0, \sigma_\epsilon^2), \quad \forall t$$

LHS	RHS	Estimate	Std. Error	t value	p	R^2
$\widehat{\text{SD}}(\widehat{y}_{t\|t-1})$	(Intercept)	0.001	<0.001	3.544	0.001	0.363
	$\widehat{\text{SD}}(\widehat{y}_{t\|t-2})$	0.403	0.097	4.135	<0.001	
$\widehat{\text{SD}}(\widehat{y}_{t\|t-2})$	(Intercept)	0.003	<0.001	6.305	<0.001	0.353
	$\widehat{\text{SD}}(\widehat{y}_{t\|t-3})$	0.352	0.072	4.905	<0.001	
$\widehat{\text{SD}}(\widehat{y}_{t\|t-3})$	(Intercept)	0.003	<0.001	9.031	<0.001	0.338
	$\widehat{\text{SD}}(\widehat{y}_{t\|t-4})$	0.420	0.051	8.183	<0.001	
$\widehat{\text{SD}}(\widehat{y}_{t\|t-4})$	(Intercept)	0.002	<0.001	3.778	0.001	0.746
	$\widehat{\text{SD}}(\widehat{y}_{t\|t-5})$	0.479	0.052	9.152	<0.001	
$\widehat{\text{SD}}(\widehat{y}_{t\|t-5})$	(Intercept)	0.003	0.001	5.359	<0.001	0.599
	$\widehat{\text{SD}}(\widehat{y}_{t\|t-6})$	0.284	0.069	4.091	<0.001	
$\widehat{\text{SD}}(\widehat{y}_{t\|t-6})$	(Intercept)	0.004	0.001	5.722	<0.001	0.731
	$\widehat{\text{SD}}(\widehat{y}_{t\|t-7})$	0.292	0.045	6.543	<0.001	

Table 2.2: Regressions of disagreement on disagreement about the same event from the previous period, and their R^2 coefficients. The specification is repeated for horizons of one to six quarters.

Inflation in the current and lagged period is observed by each forecaster, indexed over i. The signal of the current period is imprecise, with a noise component $\nu_{i,t}$, but forecasters observe lagged inflation without noise. This signal structure omits later revisions in the Fed's inflation reporting.

$$y_{i,t} = y_t + \nu_{i,t}, \quad y_{i,t-1} = y_{t-1} \tag{2.3}$$

The noise is distributed

$$\nu_{i,t} \sim iid(0, \sigma_\nu^2), \quad \forall i, t$$

Let the forecaster's estimate of y_t given his information set $\mathcal{F}_{i,t-h}$ be denoted:

$$y^*_{i,t|t-h} \equiv \text{E}[y_t|\mathcal{F}_{i,t-h}]$$

If forecasters reported $y^*_{i,t|t-h}$, this would describe a basic noisy information model. However, this model is augmented with a two-part objective function that

causes the estimate of inflation, $y^*_{i,t|t-h}$, to deviate from the reported forecast, $\widehat{y}_{i,t|t-h}$. The first part of the objective is a quadratic loss for inaccuracy:

$$L^1_{i,t,h} = (y_t - \widehat{y}_{i,t|t-h})^2$$

The second part of the objective is a quadratic loss for deviating from the forecast made in the previous period:

$$L^2_{i,t,h} = (\widehat{y}_{i,t|t-h} - \widehat{y}_{i,t|t-h-1})^2$$

However, the forecaster pays the two types of costs at different times. When he adjusts his forecast between period $t-h-1$ and period $t-h$, he pays the cost L^2 when he reports the new forecast, in time $t-h$. But when he reports an inaccurate forecast, he does not pay the cost L^1 until the realization of the variable is reported in time $t+1$ and his inaccuracy can be verified.

Suppose forecasters time-discount with an individual discount factor $0 < \beta_i \leq 1$, with weight μ_i for L^2. A forecaster in period $t-H$ that is forecasting an event in period t will seek to minimize the loss from each forecast made between now and when the event is realized:

$$L_{i,t} = \sum_{h=0}^{H} \mathrm{E}[\beta_i^{h+1}(y_t - \widehat{y}_{i,t|t-h})^2 + \beta_i^{H-h}\mu_i(\widehat{y}_{it,|t-h} - \widehat{y}_{i,t|t-h-1})^2|\mathcal{F}_{i,t-H}] \tag{2.4}$$

The first order condition for the optimal reported forecast is

$$0 = \mathrm{E}[\beta^{h+1}(y_t - \widehat{y}_{i,t|t-h}) - \beta^{H-h}\mu_i(\widehat{y}_{i,t|t-h} - \widehat{y}_{i,t|t-h-1}) + \beta^{H-h+1}\mu_i(\widehat{y}_{i,t|t-h+1} - \widehat{y}_{i,t|t-h})|\mathcal{F}_{i,t-H}] \tag{2.5}$$

The term $\widehat{y}_{i,t|t-h+1}$ can be rewritten, such that the first order condition for $\widehat{y}_{i,t|t-h}$ is in terms of $\widehat{y}_{i,t|t-h-1}$ and $y^*_{i,t|t-h}$. To do this, assume that forecasters report the truth when it is publicly observed: $\widehat{y}_{i,t|t+1} = y_t$. Then the first order condition for $h = 0$ is

$$\mathrm{E}[\widehat{y}_{i,t|t}|\mathcal{F}_{i,t-H}] = \frac{\mathrm{E}[(1+\beta^H\mu_i)y_t + \beta^{H-1}\mu_i\widehat{y}_{i,t|t-1}|\mathcal{F}_{i,t-H}]}{1+\beta^{H-1}\mu_i+\beta^H\mu_i} \tag{2.6}$$

This condition for $\widehat{y}_{i,t|t}$ can then be substituted into the analogous condition for $h = 1$. By iterating this substitution, $\widehat{y}_{i,t|t-1}$ can be written in terms of $\widehat{y}_{i,t|t-2}$ and y_t:

$$\widehat{y}_{i,t|t-h} = B_{i,h}y^*_{i,t|t-h} + (1-B_{i,h})\widehat{y}_{i,t|t-h-1} \tag{2.7}$$

In other words, when forecaster i reports expected inflation for h periods ahead, he reports a weighted average of his estimate of inflation and the forecast that he reported in the previous period.

For identification, μ_i is set as a function of β_i such that forecasters have a common weight for incorporating their true estimate of inflation in their nowcast report (i.e., some weight close to 1). That is, there exists a B such that $B = B_{i,0}\ \forall i$, implying that $\mu_i = (B-1)/(1-B(1+\beta_i^{-1}))$.

The function $B_{i,h}$ has a few salient features. It is decreasing in h and β_i. Moreover, $\widehat{\mathrm{Var}}(B_{i,h})$ of a given sample of forecasters with $\mathrm{Var}(\beta_i) > 0$ is increasing in h. These features are summarized in Figure 2.1, in which the weight $B_{i,h}$ is plotted for $B = 0.95$ and varying levels of β_i.

2.4.1 State space

The model described by Equations (1) - (3) and (7) can be written as a Markov model.

Suppose in time t that forecasters report estimates of inflation for the next H quarters. Then the state equation can be written:

$$\begin{bmatrix} x_{t+H} \\ y_{t+H} \\ y_{t+H-1} \\ \vdots \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} \phi & 0 & 0 & 0 & \dots & 0 \\ \phi & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} x_{t+H-1} \\ y_{t+H-1} \\ y_{t+H-2} \\ \vdots \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{t+H} \\ \epsilon_{t+H} + u_{t+H} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{2.8}$$

$$\boldsymbol{s}_t = \boldsymbol{F}\boldsymbol{s}_{t-1} + \boldsymbol{v}_t$$

Then $\boldsymbol{v}_t \sim iid\ N(\boldsymbol{0}, \boldsymbol{Q})$ where

$$\boldsymbol{Q} = \begin{bmatrix} \sigma_\epsilon^2 & \sigma_\epsilon^2 & 0 & \dots & 0 \\ \sigma_\epsilon^2 & \sigma_\epsilon^2 + \sigma_u^2 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

The measurement equation includes the signal of current inflation with noise, and last period's inflation without noise:

$$\begin{bmatrix} \tilde{y}_{i,t} \\ \tilde{y}_{t-1} \end{bmatrix} = \begin{bmatrix} 0 & \dots & 1 & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t+H} \\ y_{t+H} \\ y_{t+H-1} \\ \vdots \\ y_t \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \nu_{i,t} \\ 0 \end{bmatrix} \tag{2.9}$$

$$\tilde{\boldsymbol{y}}_{i,t} = \boldsymbol{H}\boldsymbol{s}_t + \boldsymbol{w}_{i,t}$$

with $\boldsymbol{w}_{i,t} \sim iid\ N(0, \boldsymbol{R})$ such that

$$\boldsymbol{R} = \begin{bmatrix} \sigma_\nu^2 & 0 \\ 0 & 0 \end{bmatrix}$$

The forecasters report a weighted average of their estimate of the state space, $s_{i,t|t-h}$, and their reported forecast from last period, $z_{i,t|t-h-1}$. Recall that

$$\widehat{y}_{i,t|t-h} = B_{i,h} y^*_{i,t|t-h} + (1 - B_{i,h})\widehat{y}_{i,t|t-h-1}$$

This condition can be rewritten in the state space. Let

$$B_i^s = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & B_{i,H-1} & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & B_{i,H-2} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}$$

And also let

$$B_i^z = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 1 - B_{i,H-1} & 0 & \dots & 0 \\ 0 & 1 - B_{i,H-2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

The condition for the vector of forecasts that forecaster i reports in time t, $z_{i,t|t}$, can be written:

$$\begin{bmatrix} \widehat{y}_{i,t+H|t} \\ \widehat{y}_{i,t+H-1|t} \\ \vdots \\ \widehat{y}_{i,t+1|t} \end{bmatrix} = B_i^s \begin{bmatrix} x_{t+H|t} \\ y_{t+H|t} \\ y_{t+H-1|t} \\ \vdots \\ y_{t|t} \\ y_{t-1} \end{bmatrix} + B_i^z \begin{bmatrix} \widehat{y}_{i,t+H-1|t-1} \\ \widehat{y}_{i,t+H-2|t-1} \\ \vdots \\ \widehat{y}_{i,t|t-1} \end{bmatrix} \tag{2.10}$$

$$\boldsymbol{z}_{i,t|t} = \boldsymbol{B_i^s s}_{i,t|t} + \boldsymbol{B_i^z z}_{i,t-1|t-1}$$

Because the system is linear and Gaussian, the optimal forecast is given by the Kalman filter.

2.4.2 Illustration of the model

In this section, I briefly illustrate the stylized facts of the model given by Equations (7) - (9) using simulated data. The following exposition is informal; this material is discussed analytically in appendices. Appendix A.1 derives that a noisy information model will not exhibit the pattern of increasing partial correlation between disagreement and the level, change, and lagged disagreement that are exhibited in Tables 2.1 and 2.2. Appendix A.2 derives the relationship between between disagreement and the level and change in inflation with heterogeneous consistency weights.

Suppose there is an inflation process with $\phi = 0.99$, $\sigma_\epsilon = 0.01$, and $\sigma_u = 0.01$. Further suppose that the signals have the variance $\sigma_\nu = 0.2$, and forecasters have a weight parameter $B = 0.99$ (these values are taken from the GMM estimates below). Figure 2.2 plots a simulated inflation process and forecasts for horizons of 1 and 8 periods. In the top two panels are the forecasts in the case where the forecasters have only an accuracy objective (a noisy information model), and in the

bottom two panels are the forecasts for the case where forecasters have both an accuracy and a consistency objective as well as varying discount rates.

Without the consistency objective, disagreement is low and constant over time; it is invariant to the level of and change in inflation. However, the consistency objective introduces time-varying disagreement. At a 1-period horizon, this disagreement is small, but the effect is magnified at the 8-period horizon.

At the long horizon, the forecaster with $\beta = 1$ places almost no weight on consistency, and his reported forecast is relatively responsive to changes in inflation and deviations from the mean of inflation. As the discount rate decreases, the forecaster becomes increasingly less responsive to new information. Disagreement is higher after a large change in inflation, and when inflation deviates from its long-run mean.

2.5 Results

The model given by Equations (7) - (9) was estimated with a 2-step GMM. It was fit to the quarterly forecasts for horizons of 0 to 4 quarters to avoid the complications of including calendar year forecasts. The details of the estimation are given in Appendix A.3. The moment conditions are derived from (1) the inflation process, (2) from the cross-sectional variance of the state estimates, and (3) from the conditions that the forecast errors $(y^*_{i,t|t-h} - y_t)$ are uncorrelated with $y^*_{i,t-h|t-h}$ and $\Delta y^*_{i,t-h|t-h}$ for each forecaster. All forecasters with fewer than 10 usable datapoints were assigned a single value of the discount rate β_i. This created 31 values of β_i. I used 3 inflation process moments, 5 cross-sectional variance moments, and $2 * 31$ correlation moments, for a total of 70 moments to estimate 36 parameters.

The estimates are given in Table 2.3. The parameters of the inflation process — ϕ, σ_ϵ and σ_u — are consistent with those in Patton and Timmermann (2010). The estimate of the noise variance, σ_ν, is roughly half of theirs; this model is explaining part of the disagreement via the consistency parameters. The estimate of B is nearly 1, meaning that forecasters place almost no weight on consistency in their nowcast. The estimates of β_i are heterogeneous and left-skewed. That

is, with a median β_i of 0.979, most of the discount rates are clustered near 1, with a tail of lower rates. The distribution of β_is is displayed with confidence intervals in Figure 2.3. The heterogeneity makes intuitive sense given that there is a mix of different kinds of forecasters in the data. Some are employed by financial services firms, some are at public institutions, and some are at universities, so one would expect their incentives to differ.

	ϕ	σ_ϵ	σ_u	σ_ν	B	μ_β	median(β_i)	σ_β	J-test p
Estimate	0.992	0.003	0.008	0.212	0.999	0.879	0.979	0.177	0.999
SE	0.028	0.051	0.054	0.001	0.015				

Table 2.3: GMM estimates of model parameters. μ_β, median(β_i) and σ_β are the mean, median and standard deviation of the discount rates.

The J-test of overidentification suggested that the model provided a good fit to the data, with a p-value of 1. Additionally, identification was checked with a MC study, and the estimates are given in Table 2.4. Simulated data were generated for 130 time periods, for 10 forecasters, and with 1000 MC repetitions. In order for the simulation to be computationally feasible, a single parameter β replaced the heterogeneous parameters β_i, and the estimation was done with 1-step GMM with an identity weight matrix. While the bias terms were generally low, the variance for ϕ was higher than the GMM estimate, suggesting that this term is more weakly identified. The variances could also be attributable to the efficiency loss from the 1-step estimation procedure and smaller sample size, which were necessary to reduce computation time.

	ϕ	σ_ϵ	σ_u	σ_ν	β	B
DGP parameter	0.999	0.031	0.010	0.217	0.863	0.995
Bias	-0.021	-0.014	0.004	-0.012	-0.002	-0.001
SD	0.082	0.009	0.008	0.035	0.012	0.006

Table 2.4: The bias and standard deviation of the parameter estimates, estimated with a Monte Carlo simulation (1-step GMM estimator)

The fit of the model was checked by simulation. Using the true inflation series, the GMM parameters were used to simulate forecaster data. The simulated data

were then used to produce the statistics generated from the SPF data in Section 2.3. Disagreement was regressed on the absolute value of the level and the change in inflation, with the regression specifications from Table 2.1. Figures 2.4 and 2.5 plot the coefficients produced by the simulated data and by the SPF data. The simulated and the SPF data produce similar coefficient sizes.

2.6 Conclusion

This chapter provides evidence that professional forecasters exhibit behavior that varies with the horizon length. I present a simple model that induces the observed patterns in disagreement over the term of a forecast. The estimates show that a significant part of disagreement between forecasters can be explained with heterogeneity in discount rates rather than signal noise. The analysis could be extended in several ways. The simplest would be to increase the horizon lengths using the proprietary Consensus Forecast survey. The inclusion of longer horizons might necessitate a more flexible parameterization of time discounting to capture curvature in the relationships between disagreement and the level of and change in inflation.

The idea of a consistency objective is related to the idea of sticky information, in that both posit a cost to changing ones estimate. Sticky information is arguably an uneasy fit for professional forecaster data, as it seems implausible that professionals would not be able to process new information every quarter. However, even if the idea of a consistency objective is abstractly plausible for professional forecasters, I have completely bracketed the question of the exact nature of this cost. It may be a reputational cost, an information constraint, a computational cost, or some other kind of adjustment cost. Further work would be required not only to disentangle these options, but to also consider whether or how this idea could be useful for modeling household expectations or firm price-setting decisions.

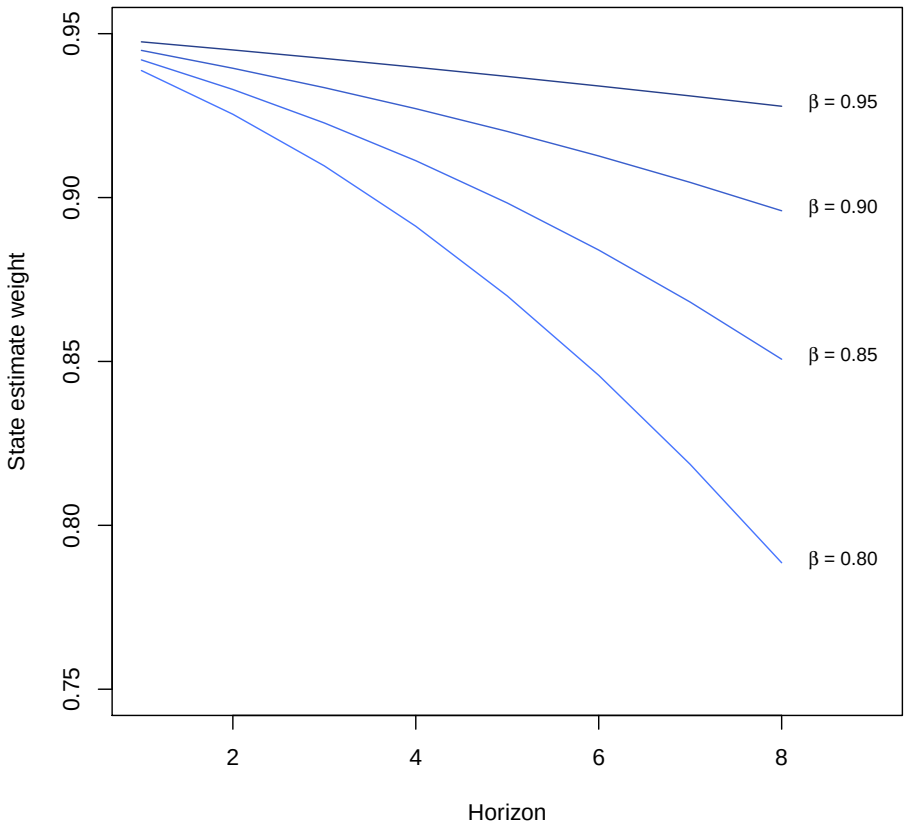


Figure 2.1: The weight $B_{i,h}$ from the optimal reported forecast condition, plotted over different horizons for different discount rates.

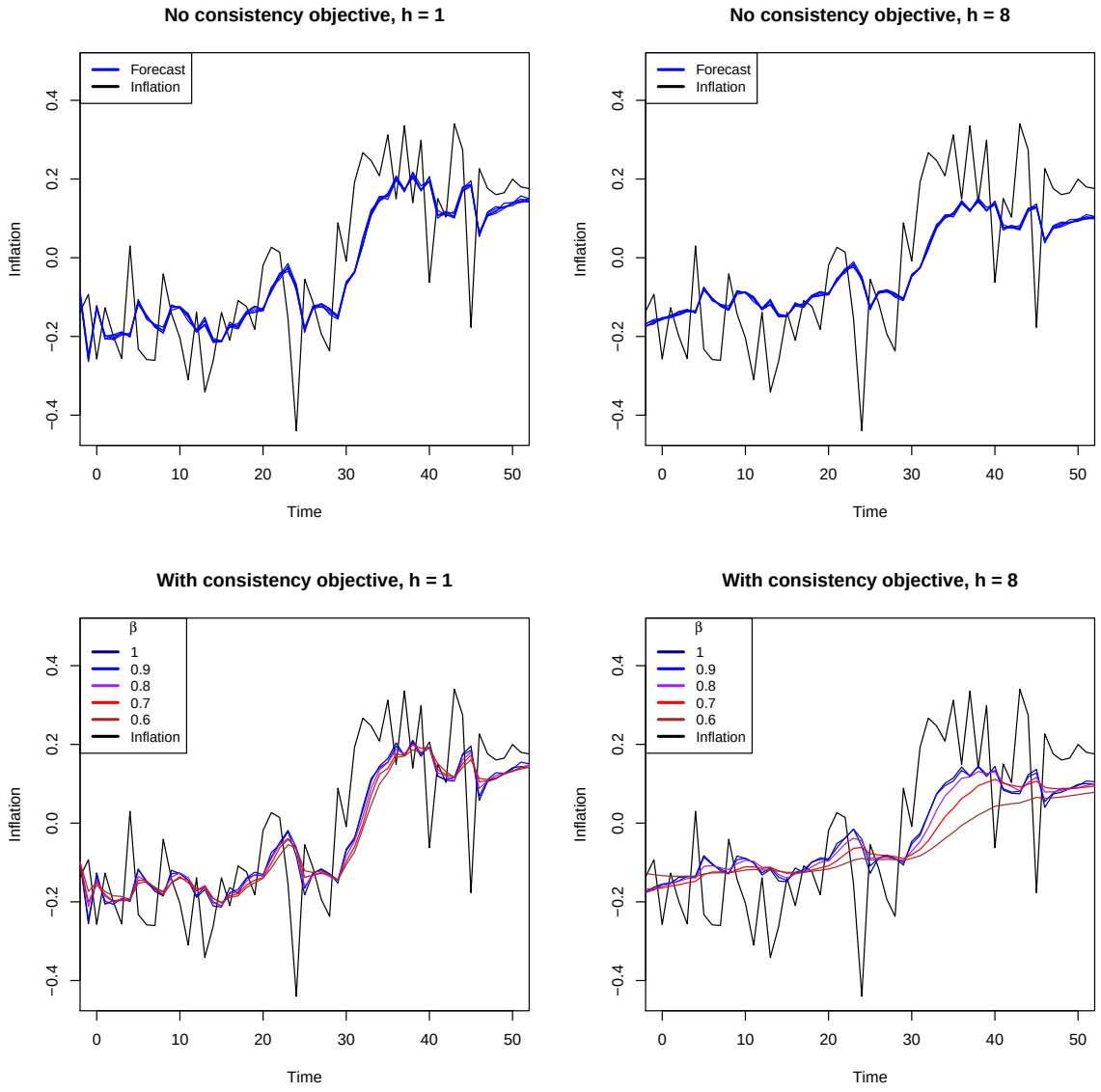


Figure 2.2: A simulated inflation process with simulated forecasts. The left panels have a horizon length of 1; the right panels have a horizon length of 8. The top panels give the forecasts that would be reported if the forecasters had no consistency objective (the Kalman filter estimates). The bottom panels show reported forecasts with a consistency objective, for forecasters with varying discount rates.

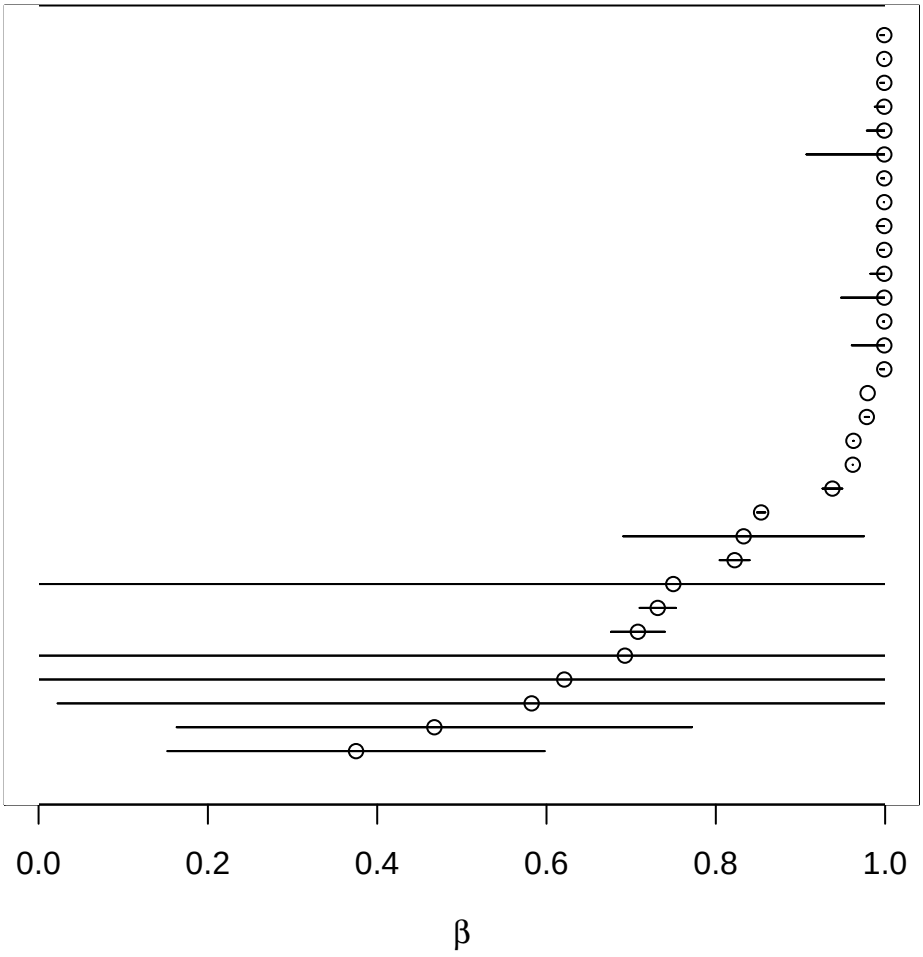


Figure 2.3: The set of β_i point estimates and confidence intervals.

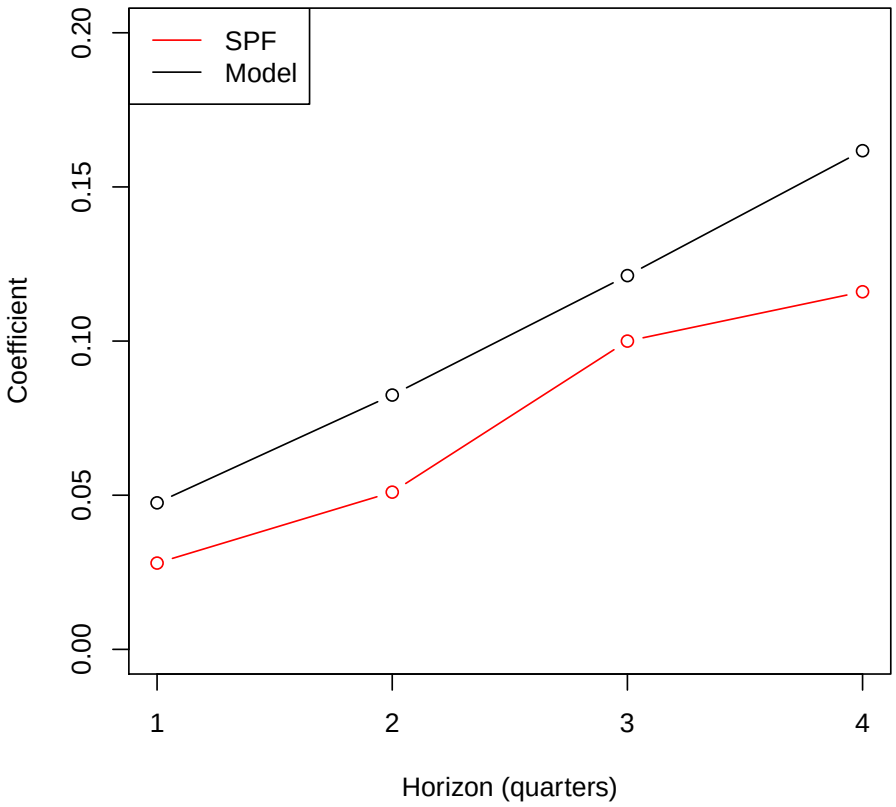


Figure 2.4: Coefficients from regressions of disagreement on the unsigned level of inflation, plotted for different horizon lengths. In red are the coefficients for the SPF data, and in black are the coefficients for the data simulated from the estimated DGP.

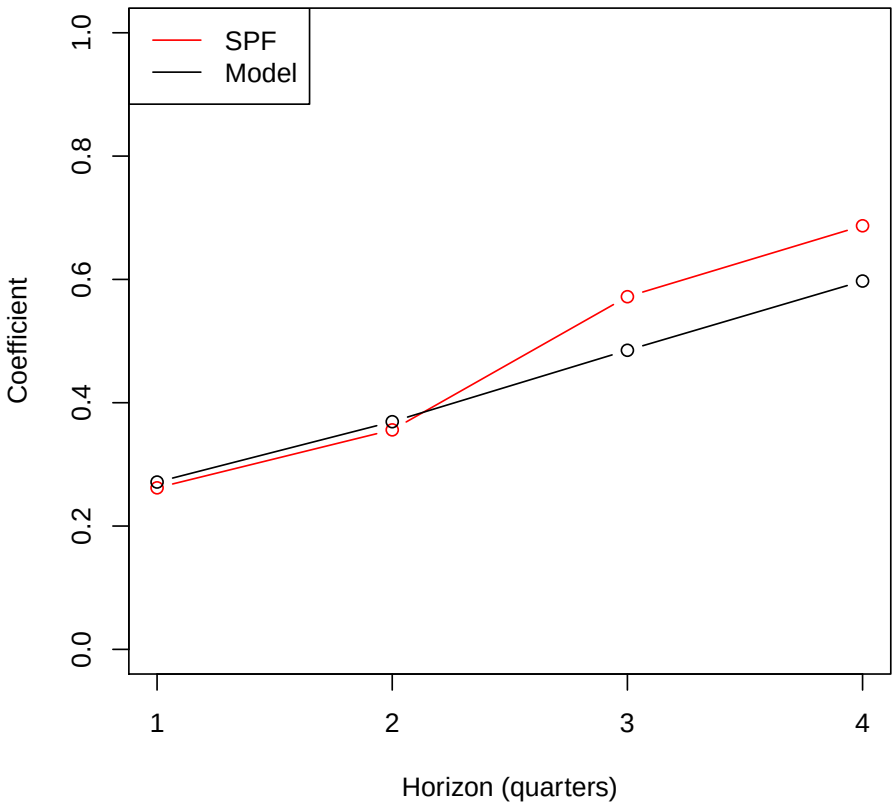


Figure 2.5: Coefficients from regressions of disagreement on the unsigned change in inflation, plotted for different horizon lengths. In red are the coefficients for the SPF data, and in black are the coefficients for the data simulated from the estimated DGP.

3

Estimation of a Rational Inattention model with multiple volatility regimes

3.1 Introduction

Since the proposal of Rational Inattention (RI) as a research agenda (Sims 1998, Sims 2003), it has been applied to a large and growing range of topics. In addition to methodological papers (Sims et al. 2005, Sims 2006, Maćkowiak et al. 2016), it has been studied in the context of monetary policy and business cycles (Adam 2007, Paciello and Wiederholt 2013, Sims et al. 2010, Maćkowiak and Wiederholt 2015), price setting (Maćkowiak et al. 2009, Maćkowiak and Wiederholt 2009), consumer search (Matějka and McKay 2012, Matějka 2015a, Matějka 2015b), as well as a variety of other applications: Maćkowiak and Wiederholt 2012 (limited liability), Matějka and Tabellini 2016 (elections), Peng and Xiong 2006 (asset pricing), Van Nieuwerburgh and Veldkamp 2009 (real estate). More recently, work has been done in the area of discrete choice (Caplin et al. 2014, Matějka et al. 2010, Matějka and McKay 2014, Steiner et al. 2015).

RI models in stationary Gaussian systems have had two features: (1) they are calibrated with constant levels of volatility and (2) they include idiosyncratic shocks to agents that are unobserved by the econometrician. In these cases, the assumption of RI behavior is required in order to estimate the attention allocated

to the idiosyncratic shock. That is, the estimation does not attempt to leverage the RI assumption for prediction. To my knowledge, the linear-quadratic Gaussian RI literature has previously sidestepped the issue of modeling the optimal signal choice by either assuming a particular signal form without proving its optimality (Paciello and Wiederholt 2013) or deriving results without solving for the optimal signal form (Sims 2003, Maćkowiak and Wiederholt 2009, Maćkowiak and Wiederholt 2015). The working paper Maćkowiak et al. (2016) solves for a dynamic RI solution to an autoregressive moving average (ARMA) process but does not estimate an RI model from data.

The goal of this chapter is to empirically investigate the model developed in Maćkowiak and Wiederholt (2009), one of the most widely cited in the RI literature. The analysis is conducted in two parts. The first part is to develop results in order to estimate agents' attention allocation in a steady-state constant volatility regime. The second part is to investigate whether the RI framework has predictive power in the following sense. I examine whether the assumption of constant attention capacity across multiple volatility regimes is supported by the data. Specifically, I estimate the steady-state model separately over two volatility regimes displays and check whether the attention capacity for each sector is correlated between regimes.

Implicitly, this analysis is investigating whether the autocovariances of the RI agent's response variable display the following pattern. When the volatility of a variable is high, RI theory predicts agents will pay greater attention to it, increasing the signal-to-noise ratio for that variable by decreasing their signal-to-noise ratio for some other variable. In other words, the RI model predicts comovement between volatility of a variable and the agent's signal-to-noise ratio for that variable, and also predicts comovement in the signal-to-noise ratio for different variables that the agent must track. I examine whether this comovement supports the hypothesis of a constant attention capacity over the length of the sample.

Using sector-level price data, I estimate the model in Maćkowiak and Wiederholt (2009) over two well-established regimes: the high-volatility period of 1976-1983, and the low-volatility period of 1992-2005. This test is related to the literature on the

volatility of prices over this period. A number of papers (Kim and Nelson (1999), McConnell and Perez-Quiros (2000), Stock and Watson (2003), among others) have noted the Great Moderation reduction in macro volatility of prices. However, evidence on the trend of microeconomic price volatility is inconclusive over this same period, see Davis and Kahn (2008) for a review. While the volatility of firm-level employment growth has fallen (Davis and Kahn, 2008), household consumption changes have not declined in volatility, and individual earnings uncertainty has risen (Cunha and Heckman, 2007). The relationship between macro- and micro-volatility remains open to interpretation.

The point that time-varying volatility sheds light on the issue of attention allocation is made in a brief section at the end of Coibion and Gorodnichenko (2015). This paper uses professional forecaster survey data (in particular, ex ante forecast revisions) to measure information rigidities by estimating the forecasters' Kalman gain. They note that their estimates of the Kalman gain among professional forecasters vary positively with volatility for the US GDP growth series. They take this evidence as supportive of an RI model. However, their evidence is limited in the sense that they only estimate the Kalman gain for one series over time. They do not present evidence that forecasters are changing attention allocation between variables. Consequently their empirics are equally supportive of the hypothesis that there is an exogenous increase in the amount of information about GDP growth in periods of high volatility.

My estimates suggest an increase in attention allocation particularly to idiosyncratic conditions in the second regime. It is plausible that costs to gathering individual-specific information have been reduced by information technologies, whereas the availability or cost of collecting macro information is relatively constant over the sample. At the sector level, attention capacity in the first regime is not strongly predictive of attention capacity in the second regime. The estimates have a number of significant caveats that I discuss with the results.

In Section 3.2, I provide an introduction to the information theory. In Section 3.3, I outline the particular model that I test, Maćkowiak and Wiederholt (2009),

henceforth MW. In Section 3.4, I discuss identification of MW. In Section 3.5, I discuss the use of volatility regimes. In Section 3.6, I provide the empirical results. Section 3.7 concludes.

3.2 Information theory

Rational Inattention theory posits that agents receive information transmitted through a channel with limited information flow, such that input data exit the channel with error.

Suppose the input is a random variable X with pdf f_X. The channel outputs a random variable S with pdf $f_{S|X}(s|x)$. Then we can construct the conditional pdf for $X|S$ as:

$$f_{X|S}(x|s) = \frac{f_{X,S}(x,s)}{f_S(s)} = \frac{f_X(x)f_{S|X}(s|x)}{\int f_{X,S}(x,z)dx} \tag{3.1}$$

Entropy is defined as $H(X) = \mathrm{E}(-\log(X))$. The information acquired about X from observing S is the change in entropy

$$I(X;S) = H(X) - H(X|S) = \mathrm{E}(-\log(f_X(x))) - \mathrm{E}(-\log(f_{X|S}(x|s))|S) \tag{3.2}$$

and is called the mutual information.

We can model limited attention as a bound on mutual information

$$I(X;S) \leq \kappa \tag{3.3}$$

The entropy $H(X)$ of a normally distributed random variable X is

$$H(X) = \frac{1}{2}\log(2\pi e\sigma_X^2) \tag{3.4}$$

where σ_X^2 denotes the variance of X and e is the natural number. The conditional entropy of X given S, when X and S have a joint multivariate normal distribution, equals

$$H(X|S) = \frac{1}{2}\log(2\pi e\sigma^2_{X|S}) \tag{3.5}$$

where $\sigma^2_{X|S}$ denotes the conditional variance of X given S.

The constraint on information flow implies that

$$\frac{\sigma^2_X}{\sigma^2_{X|S}} \leq \exp(2\kappa) \tag{3.6}$$

An agent is constrained in her choice of signal S. The agent cannot choose a signal that is too informative, i.e. a signal for which $\sigma^2_{X|S}$ is too small.

In the multivariate normal case, this condition can be written as

$$\frac{\det \Sigma_X}{\det \Sigma_{X|S}} \leq \exp(2\kappa) \tag{3.7}$$

where Σ_X and $\Sigma_{X|S}$ are prior and posterior positive definite covariance matrices.

The above hints at how changes in volatility can have testable implications for the RI hypothesis. In the multivariate case, only the determinant of the posterior variance matrix $\det \Sigma_{X|S}$ is pinned down by the constraint. This leaves degrees of freedom in $\Sigma_{X|S}$ over which the agent must optimize. If the model is sufficiently flexible, then for a given covariance matrix Σ_X, there will be a parameterization for which the agent's choice of $\Sigma_{X|S}$ is optimal. However, if we observe multiple volatility regimes, then the capacity may have to differ between these regimes to rationalize the agent's choices.

3.3 Motivating model

The discussion of identification is motivated by the model in MW. In this section, I outline the MW model specification. This outline is very brief, with the aim of providing sufficient detail to contextualize a model in which the agent must attend to two independent stationary Gaussian processes, an aggregate and an idiosyncratic series of shocks. Please see MW for technical details. While the focus on two shocks is restrictive, especially for an empirical exercise, the idiosyncratic shocks can be understood as a convolution of multiple shocks.

Suppose there is a continuum of firms indexed by $i \in [0,1]$. Each firm i sells good i. The firm sets the price of its good, P_{it} to maximize the expected discounted sum of profits:

$$P_{it}^* = \arg\max_{\{P_{it}\}} E_{it}\left[\sum_{\tau=t}^{\infty} \beta^{\tau-t}\pi(P_{i\tau}, P_\tau, Y_\tau, Z_{i\tau})\right] \tag{3.8}$$

The profit function π is a function of the price set by the firm P_{it}, the price level P_t, real aggregate demand Y_t, and idiosyncratic firm-specific conditions Z_{it}. The following assumes that the function π is twice continuously differentiable, homogenous of degree zero in its first two arguments, and single-peaked in P_{it} for given P_t, Y_t, and Z_{it}. Suppose that $Q_t = P_t Y_t$, where Q_t is nominal demand, an exogenous process, and let the price level be $\log P_t = \int_0^1 \log P_{it} di$, as in a model of monopolistic competition.

The model uses a log-linearized approximation around the nonstochastic solution. The 2nd-order Taylor approximation is (see MW Equation 18 for derivations):

$$\tilde{\pi}(p_{it}, p_t, y_t, z_{it}) = \hat{\pi}_1 p_{it} + \frac{\hat{\pi}_{11}}{2} p_{it}^2 + \hat{\pi}_{12} p_{it} p_t + \hat{\pi}_{13} p_{it} y_t \tag{3.9}$$

$$+ \hat{\pi}_{14} p_{it} z_{it} + \text{terms independent of } p_{it} \tag{3.10}$$

Where p_{it}, p_t, y_t, z_{it} are log-deviations from steady-state of P_{it}, P_t, Y_t, Z_{it} respectively. $\hat{\pi}_1, \hat{\pi}_{11}, \hat{\pi}_{12}, \hat{\pi}_{13}, \hat{\pi}_{14}$ are constants. We can decompose the optimal price into a aggregate and idiosyncratic terms:

$$p_{it}^* = \Delta_t + \alpha_z z_{it} \tag{3.11}$$

where Δ_t are aggregate conditions, z_{it} are idiosyncratic conditions, and α_z is a constant. There is no general closed-form solution for Δ_t. As the innovations to q_t are the only exogenous aggregate component, Δ_t will be a function of them:

$$\Delta_t = \sum_{l=0}^{\infty} \rho_{\Delta,l} q_{t-l} \tag{3.12}$$

More specifically, MW derive the aggregate conditions as a function of p_t and q_t:

$$\Delta_t = (1 - \pi_p)p_t + \pi_q q_t \tag{3.13}$$

where π_p, π_q are constants. The firm's loss due to suboptimal pricing is quadratic:

$$\tilde{\pi}(p^*_{it}, p_t, y_t, z_{it}) - \tilde{\pi}(p_{it}, p_t, y_t, z_{it}) = (p^*_{it} - p_{it})^2 \tag{3.14}$$

If the firm had no information constraint, then it would set the optimal price in every period, $p_{it} = p^*_{it}$. The RI assumption is that there is a bound on the mutual information between the signal s_{it} received by the firm and the joint distribution of the aggregate and idiosyncratic conditions:

$$I(\{P_t, Z_{it}\}; s_{it}) \leq \kappa \tag{3.15}$$

The firm observes the signal and sets the price to be the expected optimal price given their information set at time t:

$$p_{it} = \mathrm{E}(p^*_{it}|\mathcal{I}_{it}),\ \mathcal{I}_{it} = \{s_{i,-\infty}, ..., s_{it}\} \tag{3.16}$$

Their objective is to choose the signal to minimize their expected loss stream:

$$\min_{s_{it}} E\left[\sum_{t=1} \beta^t (p^*_{it} - p_{it})^2\right] \tag{3.17}$$

Lastly, we assume some orthogonality conditions. Suppose that $\mathrm{E}(z_{it}) = \mathrm{E}(z_{it}z_{jt}) = 0\ \forall i \neq j$ (zero mean and pairwise independence). Further suppose that z_{it} and Δ_t are independent. From the result in Uhlig (1996), it follows that

$$\int_0^1 z_{it} di = 0 \tag{3.18}$$

3.4 Optimal signal choice

Prior to estimation, it is necessary to solve the agent's optimal signal problem. In the full model, the agent must track two independent Gaussian stationary variables over multiple regimes. In this section I build up to this model in stages. First, I discuss the case of an agent tracking a univariate stationary Gaussian process with constant volatility. Second, I discuss the case when the agent must track two independent stationary Gaussian processes with constant volatility. Third, I discuss the agent's signal choice when there is a volatility regime change.

3.4.1 Univariate optimal signal

In this section, I outline the agent's optimal steady state signal for a univariate stationary Gaussian system. The material through (3.25) is standard. It is for example used in Maćkowiak et al. (2016). The material that follows is original. It develops the result that the agent minimizes her loss by only paying attention to the most recent innovation.

Suppose that the agent is tracking a stationary Gaussian variable z_t, which can be written as an infinite-order MA process:

$$z_t = \sum_{l=0}^{\infty} \alpha_{z,l} \epsilon_{z,t-l} \tag{3.19}$$

where $\epsilon_{z,t} \sim N(0, \phi_z^2)$ and $\sum_{l=0}^{\infty} |\alpha_{z,l}| < \infty$.

Some additional notation is required to discuss the optimal signal choice. The agent's signal is denoted $\hat{z}_t$. Her information set is the entire history of signals: $\mathcal{I}_t = \{\hat{z}_t, \hat{z}_{t-1}, ..., \hat{z}_{-\infty}\}$. Every signal is orthogonal to future information: $\mathrm{E}(\hat{z}_t \epsilon_{z,t+k}) = 0, \forall k > 0$. After observing a signal at time t, the agent forms an expectation of z_t: $z_{t|t} \equiv \mathrm{E}(z_t|\mathcal{I}_t)$. This estimate has a conditional variance denoted $\sigma^2_{z,t|t} \equiv \mathrm{Var}(z_t|\mathcal{I}_t)$.

Given this notation, we can consider the set of possible signals that the agent may choose. It is useful to distinguish between the part of the signal that is a deterministic function of the history of z_t, and the part of the signal that is noise (stochastic

conditional on the history of z_t). Let $\tilde{z}_t$ be the deterministic vector component, which in this univariate example, is any function h_z of the history of the innovations:

$$\tilde{z}_t = h_z(\epsilon_{z,t}, \epsilon_{z,t-1}, ..., \epsilon_{z,-\infty}) \tag{3.20}$$

Now define the signal as a vector that is a function of the deterministic component z_t and observation noise $e_{z,t}$:

$$\hat{z}_t = g_z(\tilde{z}_t, e_{z,t}) \tag{3.21}$$

The purpose of defining the signal in this way is because this allows us to compactly write the information constraint and later results. The agents have a bound on the mutual information between $\tilde{z}_t$ and the signal $\hat{z}_t$:

$$I(\tilde{z}_t; \hat{z}_t) \leq \kappa \tag{3.22}$$

The agent minimizes her loss stream ℓ_t, discounted by a factor β. This can be written:

$$\ell_t = \min_{\{\tilde{z}_t\}} \sum_{s=t}^{\infty} \beta^{s-t} \sigma^2_{z,s|s} \text{ s.t. } I(\tilde{z}_t; \hat{z}_t) \leq \kappa \tag{3.23}$$

A standard result from information theory is that an agent minimizing quadratic loss will choose a signal such that her posterior is Gaussian and unbiased. This theorem is written in full, for example, in Cover and Thomas (2012), p 255, Theorem 8.6.6. The agent's response is:

$$z_{t|t} \sim N(z_t, \sigma^2_{t|t}) \tag{3.24}$$

Under the properties of closure of the Gaussian distribution, the signal $\hat{z}_t$ and the process z_t will be distributed jointly multivariate Gaussian. The signal is a linear combination of the innovations of z_t and a measurement error $e_{z,t}$:

$$\hat{z}_t = \sum_{l=0}^{\infty} \gamma_{z,l} \epsilon_{z,t-l} + e_{z,t} \tag{3.25}$$

The measurement error is orthogonal to the innovations: $\mathrm{E}(e_{z,t}\epsilon_{z,t-l}) = 0,\ \forall l \geq 0$.

I require one further assumption in order to later prove that agents will converge on a simple steady-state signal. This is an assumption of strictly decreasing MA coefficients for z_t, as in (3.19). That is, suppose that

$$|\alpha_{z,l}| > |\alpha_{z,l+1}| \ \ \forall\, l \geq 0 \tag{3.26}$$

This assumption of decreasing coefficients is specifically required to show that, from some arbitrary initial conditions, agents will converge within a finite number of periods on their steady-state attention allocation (see proof of Proposition 3.3 in Appendix B.3).

With this notation, we can determine the values of the coefficients $\{\gamma_{z,l}\}$, that is, which innovations the agent will pay attention to. The following propositions develop the fact that the agent will choose a scalar signal of the most recent innovation. Note that Proposition 3.1 does not require the assumption of decreasing MA coefficients from (3.26).

Proposition 3.1 *Suppose an RI agent must report an estimate $\hat{x}$ of a variable x. x is a known linear combination of jointly Gaussian variables, x_1 and x_2. Suppose that the loss function ℓ is quadratic with some diagonal positive weighting matrix W: $\ell \equiv -E((x-\hat{x})'W(x-\hat{x}))$. Further suppose that x_1 and x_2 are independent in the agent's prior distribution:*

$$(x_1, x_2)' \sim N((\mu_1, \mu_2)', diag(\sigma_1^2, \sigma_2^2)). \tag{3.27}$$

Consider an arbitrary 2×1 vector signal $\hat{x}_{1;2}$ in which both entries have nonzero covariance with both x_1 and x_2. Then there will exist a 2×1 vector signal, $(\hat{x}_1, \hat{x}_2)$, in which $Cov(\hat{x}_1, \hat{x}_2) = 0$, such that the agent can achieve the same loss with $(\hat{x}_1, \hat{x}_2)$ as with $\hat{x}_{1;2}$.

Proof: See Appendix B.1. ∎

This result says that it is optimal for the agent to preserve the independence of prior variables in the posterior when these variables must be used for multiple

forecasts. There may be other signals, that do not preserve independence, that achieve this same loss. For the remainder of the chapter, I assume that the agent chooses independence-preserving signals. The intuition for this result is clearer in the next proposition, where I begin to apply this result to the MW model. Note that Proposition 3.2 also does not require the assumption of decreasing MA coefficients from (3.26).

Proposition 3.2 *Consider an RI agent tracking z_t as described in (3.19) - (3.23). The agent will choose to receive independent signals of each innovation $\epsilon_{z,t-l}$, $l \geq 0$. That is, $\tilde{z}_t = (\epsilon_{z,t}, \epsilon_{z,t-1}, ..., \epsilon_{z,-\infty})'$.*

Proof: See Appendix B.2. ∎

The intuition for this result is as follows. If it were the case that the model were static, and the agent only had to report an estimate $\hat{z}_t$ for a single period, then it would be optimal for the agent to receive a signal of precisely the quantity reported, ie they would choose the signal $\tilde{z}_t = z_t$. However, the model is dynamic, and the agent will use the information gathered in period t in future periods to predict z_{t+1}, z_{t+2}, etc. Therefore, if z_t is an arbitrary infinite-order MA process, the agent derives utility from the accurate prediction of infinite linear combinations of the innovations. There will therefore exist an optimal signal for the agent, in which she receives independent signals of each innovation. In each subsequent period, the innovation signals can be arbitrarily recombined for each nowcast.

We can now show that the agent will converge on steady-state behavior, in which she only pays attention to the most recent innovation. Up until this point, we have not used the condition from (3.26) that the MA coefficients for z_t are decreasing. However, the following proposition makes use of this assumption.

Proposition 3.3 *Consider an RI agent tracking z_t as described in (3.19) - (3.23). In particular, suppose that z_t has strictly decreasing MA coefficients: $|\alpha_{z,l}| > |\alpha_{z,l+1}|$ $\forall l \geq 0$. Suppose there are initial conditions such that in period t it is optimal for the agent to pay non-zero attention to (at least one) lagged innovation $\epsilon_{z,t-k}$:*

$$\frac{Var(\epsilon_{z,t-k}|\mathcal{I}_{t-1})}{Var(\epsilon_{z,t-k}|\mathcal{I}_t)} > 1,\ k > 0 \tag{3.28}$$

Within a finite number of periods $j^* < \infty$, the agent will pay zero attention to all lagged innovations:

$$\frac{Var(\epsilon_{z,t+j-k}|\mathcal{I}_{t+j-1})}{Var(\epsilon_{z,t+j-k}|\mathcal{I}_{t+j})} = 1\ \forall j > j^*, k > 0 \tag{3.29}$$

Proof: See Appendix B.3. ∎

The intuition for this result is as follows. In order for agents to converge from some arbitrary initial conditions to the steady-state behavior of only paying attention to the current innovation, it must be the case that the marginal utility of paying attention the the current innovation is higher than the marginal utility of paying attention to any lagged innovation. One way of insuring that this will be the case is to require that the MA coefficients on the innovations are strictly decreasing. Then, regardless of the level of time-discounting in β, information about recent innovations will be more valuable to agents than information about lagged innovations. Given arbitrary initial conditions, only learning about the current innovation will be the absorbing state.

In the next proposition, I prove that paying attention to the most recent innovation is not only the absorbing strategy, but it is also the loss-minimizing strategy.

Proposition 3.4 *Consider an RI agent as described in (3.19) - (3.23). Consider the limiting loss associated with a signal strategy, $\ell_t(\hat{z}_t)|\mathcal{I}_t$ as $t \to \infty$. The agent will minimize this loss by always choosing a scalar signal of the current innovation, ie $\tilde{z}_t = \epsilon_{z,t}$:*

$$\arg\left(\min_{\tilde{z}_t}\left(\lim_{t\to\infty}\sum_{s=t}^{\infty}\beta^{s-t}\sigma^2_{s|s}\right)\right) = \epsilon_{z,t} \tag{3.30}$$

Proof: See Appendix B.4. ∎

Proposition 3.4 does *not* make use of the assumption of decreasing MA coefficients from (3.26). However, in the absence of this assumption, we cannot insure that

agents will ever converge to this loss-minimizing behavior. Insuring that the behavior will converge is necessary particularly because the analysis will examine multiple volatility regimes, and in the transition between regimes, agents will not be in steady-state behavior.

This completes the specification of the optimal signal choice for the agent in a univariate setting with constant volatility.

To show the identification result for this section, we must find the form of the response $z_{t|t}$. In order to do this, I add the additional condition that the agent uses Kalman filtering. This assumption, which is used, for instance, in Maćkowiak et al. (2016), is justifiable in the sense that the Kalman filter is optimal in this setting.

The following proposition gives the distribution of $z_{t|t}$.

Proposition 3.5 *Consider an RI agent as described in (3.19) - (3.23) who uses Kalman filtering. Suppose, following Proposition 3.3, that she pays attention only to ϵ_{zt} in period t, $\forall t$. She tracks the variable*

$$z_t = \sum_{l=0}^{\infty} \alpha_{z,l} \epsilon_{z,t-l} \tag{3.31}$$

where $\epsilon_{z,t} \sim N(0, \phi_z^2)$. Then the agent's response function $z_{t|t}$ will have the following form:

$$z_{t|t} = \sum_{l=0}^{\infty} \alpha_{z,l} \epsilon_{z,t-l|t-l} \tag{3.32}$$

where $\epsilon_{z,t|t} \sim N(0, \phi_z^2)$ and

$$Cor(z_t, z_{t|t}) = \frac{1}{1 + (\exp(2\kappa_z) - 1)^{-1}} \tag{3.33}$$

Proof: See Appendix B.5. ∎

The point of this result is that the attention parameter is not identified from the response variable alone. If the econometrician observes the response $z_{t|t}$ but not the fundamental shock series z_t, then the parameters of z_t are identified. The econometrician will have consistent estimators of all MA coefficients $\{\alpha_{z,l}\}$ and

also the residual variance ϕ_z^2. However, this will not identify κ_z. To identify the attention, an estimate of $\text{Cor}(z_t, z_{t|t})$ is required.

3.4.2 Bivariate optimal signal

This section follows the setup given in MW. Suppose that the agent is now tracking a variable p_t^* that is the sum of two independent stationary Gaussian variables with constant volatility. The specification is:

$$p_t^* = \Delta_t + z_t \tag{3.34}$$

$$\Delta_t = \sum_{l=0}^{\infty} \alpha_{\Delta,l} \epsilon_{\Delta,t-l} \tag{3.35}$$

$$z_t = \sum_{l=0}^{\infty} \alpha_{z,l} \epsilon_{z,t-l} \tag{3.36}$$

where, for $i \in \{\Delta, z\}$, $\epsilon_{i,t} \sim N(0, \phi_i^2)$ and $\sum_{l=0}^{\infty} |\alpha_{i,l}| < \infty$. Further suppose that $|\alpha_{i,l}| > |\alpha_{i,l+1}| \; \forall l \geq 0$. This assumption of decreasing MA coefficients is again required to prove that agents' signal choice converges to paying attention to the current innovation (see Proposition 3.6).

Following MW, I assume that the agent separately tracks Δ_t and z_t, which captures the idea that paying attention to aggregate and idiosyncratic conditions are separate activities. Let $\kappa_{\Delta t}$ and κ_{zt} denote the attention allocated at time t to Δ_t and z_t, respectively.

Agents choose the attention allocation to minimize the loss subject to the information constraint. If the information constraint binds, then the following holds:

$$\min_{\hat{z}_t, \hat{\Delta}_t} \sum_{s=t}^{\infty} \beta^{s-t} \sigma_{s|s}^2, \text{ where } \sigma_{t|t}^2 = \sigma_{\Delta,t|t}^2 + \sigma_{z,t|t}^2 \text{ s.t.} \tag{3.37}$$

$$\frac{\text{Var}(\tilde{\Delta}_t | \mathcal{I}_{t-1})}{\text{Var}(\tilde{\Delta}_t | \mathcal{I}_t)} \frac{\text{Var}(\tilde{z}_t | \mathcal{I}_{t-1})}{\text{Var}(\tilde{z}_t | \mathcal{I}_t)} = \exp(2(\kappa_{\Delta t} + \kappa_{zt})) = \exp(2\kappa) \tag{3.38}$$

Consider a steady-state allocation in which $\kappa_{\Delta_t} = \kappa_\Delta$ and $\kappa_{z_t} = \kappa_z$ for all t. Given Proposition 3.4, all attention will be devoted to the current innovations $\epsilon_{\Delta,t}$ and $\epsilon_{z,t}$. The marginal utility of paying attention to $\epsilon_{\Delta,t}$ is:

$$\frac{\partial \ell_t}{\partial \kappa_\Delta} = \sum_{s=t}^{\infty} \beta^{s-t} \frac{\partial \sigma^2_{s|s}}{\partial \kappa_\Delta} = -2 \sum_{s=t}^{\infty} \beta^{s-t} \alpha^2_{s-t} \phi^2_\Delta \exp(-2\kappa_\Delta) \tag{3.39}$$

The analogous condition holds for z. Note the following concavity:

$$\frac{\partial^2 \ell_t}{\partial \kappa^2_\Delta} < 0, \ \frac{\partial^2 \ell_t}{\partial \kappa^2_z} < 0, \ \frac{\partial^2 \ell_t}{\partial \kappa_\Delta \partial \kappa_z} = 0 \tag{3.40}$$

Consequently, if an interior solution exists (in which nonzero attention is paid to both Δ_t and z_t), then the following boundary conditions hold:

$$\frac{\partial \ell_t}{\partial \kappa_\Delta} < \frac{\partial \ell_t}{\partial \kappa_z}\bigg|_{(\kappa_\Delta, \kappa_z)=(\kappa, 0)} \text{ and } \frac{\partial \ell_t}{\partial \kappa_\Delta} > \frac{\partial \ell_t}{\partial \kappa_z}\bigg|_{(\kappa_\Delta, \kappa_z)=(0, \kappa)} \tag{3.41}$$

In an interior solution,

$$\frac{\partial \ell_t}{\partial \kappa_\Delta} = \frac{\partial \ell_t}{\partial \kappa_z} \text{ and } \kappa_\Delta + \kappa_z = \kappa \tag{3.42}$$

Inserting the result from (3.39), the attention allocation solves the system

$$\sum_{s=t}^{\infty} \beta^{s-t} \alpha^2_{s-t} \phi^2_\Delta \exp(-2\kappa_\Delta) = \sum_{s=t}^{\infty} \beta^{s-t} \alpha^2_{s-t} \phi^2_z \exp(-2\kappa_z) \text{ and} \tag{3.43}$$

$$\kappa_\Delta + \kappa_z = \kappa \tag{3.44}$$

If the interior solution does not exist, then either $\kappa_\Delta = \kappa$ or $\kappa_z = \kappa$.

The following two propositions are corollaries of Propositions (3.3) and (3.4), now for a bivariate system.

Proposition 3.6 *Consider an RI agent as described in (3.34) - (3.38). Suppose there are initial conditions such that in period t it is optimal for the agent to pay non-zero attention to (at least one) lagged innovation $\epsilon_{i,t-k}$:*

$$\frac{Var(\epsilon_{i,t-k}|\mathcal{I}_{t-1})}{Var(\epsilon_{i,t-k}|\mathcal{I}_t)} > 1, \ k > 0, \ i \in \{\Delta, z\} \tag{3.45}$$

Within a finite number of periods $j^* < \infty$, the agent will pay zero attention to all lagged innovations:

$$\frac{Var(\epsilon_{i,t+j-k}|\mathcal{I}_{t+j-1})}{Var(\epsilon_{z,t+j-k}|\mathcal{I}_{t+j})} = 1 \; \forall j > j^*, k > 0, i \in \{\Delta, z\} \tag{3.46}$$

Proof: See Appendix B.6. ∎

Proposition 3.7 *Consider an RI agent as described in (3.34) - (3.38). Consider the limiting loss of a signal strategy, $\ell_t|\mathcal{I}_t$ as $t \to \infty$. The agent will minimize this loss by always choosing a scalar signal of the current innovations:*

$$\arg\left(\min_{\tilde{z}_t}\left(\lim_{t\to\infty}\sum_{s=t}^{\infty}\beta^{s-t}\sigma^2_{s|s}\right)\right) = (\epsilon_{\Delta,t}, \epsilon_{z,t})' \tag{3.47}$$

Proof: See Appendix B.7. ∎

Now we may turn to the issue of identification. This system differs from the univariate case in the previous section because now we assume that a variable that the agent is tracking, Δ_t, is observable to the econometrician. We can identify the correlation between Δ_t and p_t, which in turn identifies the attention. This is captured in the following proposition.

Proposition 3.8 *Suppose we observe stationary Gaussian data that consist of the following parameters from p_t and Δ_t:*

$$p_t = \sum_{l=0}^{\infty}\alpha_{p,l}\epsilon_{p,t-l}, \; \epsilon_{p,t} \sim N(0, \phi_p^2) \tag{3.48}$$

$$\Delta_t = \sum_{l=0}^{\infty}\alpha_{\Delta,l}\epsilon_{\Delta,t-l}, \; \epsilon_{p,t} \sim N(0, \phi_\Delta^2) \tag{3.49}$$

$$\rho = Cor(p_t, \Delta_t) > 0 \tag{3.50}$$

Then there exist unique attention parameters κ_Δ and κ_z that would produce the given parameters from the RI specification in (3.34) - (3.38).

If $\rho = Cor(p_t, \Delta_t) = 0$, then $\kappa_z = \kappa$ and κ is only set-identified. That is, it is possible to find a finite $\bar{\kappa}$ such that $0 \leq \kappa < \bar{\kappa}$.

Proof: See Appendix B.8. ∎

If $\rho = \text{Cor}(p_t, \Delta_t) < 0$, then it is possible to identify κ_Δ and κ_z only by respecifying the model. For instance, by rewriting the optimal price as $p_t^* = -\Delta_t + z_t$. I discuss this issue further in Appendix B.8.

The implication of this result is that data from a standard signal extraction problem can be 'rationalized'. Given arbitrary stationary Gaussian processes for p_t and Δ_t, it will be possible to find parameters for z_t, κ_Δ and κ_z that satisfy the optimality condition. Note that this will even be true if Δ_t and p_t are nearly independent. In this case, the estimated κ will be very close to 0. This is a key just-identification result.

In the next section, I discuss a non-stationary system with multiple volatility regimes. With multiple regimes, it will be possible to overidentify κ. Hence it will also be possible to examine whether the attention capacity estimated in one regime is predictive of the attention capacity in another regime.

3.5 Volatility regimes

In this section I relax the assumption of a stationary system and introduce multiple volatility regimes. Suppose in the sample of T periods, there is a structural break at $t = t^*$, such that the innovation variances change at this point. The specification of z_t is now:

$$z_t = \begin{cases} \sum_{l=0}^{\infty} \alpha_{1zl} \epsilon_{z,t-l}, & t < t^* \\ \sum_{l=0}^{\infty} \alpha_{2zl} \epsilon_{z,t-l} & t \geq t^* \end{cases} \tag{3.51}$$

$$\epsilon_{z,t} \sim \begin{cases} N(0, \phi_{1z}^2), & t < t^* \\ N(0, \phi_{2z}^2), & t \geq t^* \end{cases} \tag{3.52}$$

where $\sum_{l=0}^{\infty}, |\alpha_{1zl}| < \infty$ and $\sum_{l=0}^{\infty}, |\alpha_{2zl}| < \infty$. Additionally, $|\alpha_{1zl}| > |\alpha_{1,z,l+1}|$ and $|\alpha_{2zl}| > |\alpha_{2,z,l+1}| \ \ \forall l$.

We also need to give a distribution to Δ_t, the aggregate component. In the MW model, Δ_t is driven by innovations to q_t. We can write the regimes in terms of the innovation variance of q_t:

$$q_t = \begin{cases} \sum_{l=0}^{\infty} \alpha_{1ql} \epsilon_{q,t-l}, & t < t^* \\ \sum_{l=0}^{\infty} \alpha_{2ql} \epsilon_{q,t-l} & t \geq t^* \end{cases} \tag{3.53}$$

$$\epsilon_{q,t} \sim \begin{cases} N(0, \phi_{1q}^2), & t < t^* \\ N(0, \phi_{2q}^2), & t \geq t^* \end{cases} \tag{3.54}$$

where the MA coefficients are absolutely summable and monotonically decreasing in the lags.

We can write $\Delta_t = p_t + \frac{\hat{\pi}_{13}}{|\hat{\pi}_{11}|}(q_t - p_t)$ (MW Equation 20). Since Δ_t is a function of the endogenous variable p_t, in general it has no closed form solution. Δ_t is endogenous in the sense that the volatility of idiosyncratic demand for one agent affects the volatility of the aggregate conditions for all agents. If a mass of firms experience changes in the volatility of their idiosyncratic conditions, then they will shift attention devoted to Δ_t, which will affect p_t, which will change the volatility of Δ_t. I assume that Δ converges on a stationary distribution after a period of adjustment. That is, the following is a reasonable approximation of the distribution for some $t^{**} > t^*$:

$$\Delta_t = \begin{cases} \sum_{l=0}^{\infty} \alpha_{1\Delta l} \epsilon_{\Delta,t-l}, & t < t^* \\ \sum_{l=0}^{\infty} \alpha_{2\Delta l} \epsilon_{\Delta,t-l}, & t \geq t^{**} \end{cases} \tag{3.55}$$

$$\epsilon_{\Delta,t} \sim \begin{cases} N(0, \phi_{1\Delta}^2), & t < t^* \\ N(0, \phi_{2\Delta}^2), & t \geq t^{**} \end{cases} \tag{3.56}$$

where $\sum_{l=0}^{\infty}, |\alpha_{k,\Delta,l}| < \infty$ and $|\alpha_{k,\Delta,l}| > |\alpha_{k,\Delta,l+1}| \ \forall l$.

By Proposition 3.6, after a finite number of periods following the regime switch at t^*, the agent will only pay attention to the current innovations again. Consequently, we can use this steady-state approximation of the agent's behavior to estimate the model in each regime.

3.6 Results

3.6.1 Data

I estimate the model using data from Boivin et al. (2009). They provide a disaggregated US Personal Consumption Expenditure (PCE) index, including 89

price series on civilian durable and nondurable goods. I take each goods sector to be a price series p_{it}, and so the estimation treats each firm in the model as one sector. The list of these series is given in Appendix B.9. The data are monthly from January 1976 to June 2005, and I aggregate each to the quarterly level, as the aggregate data are only available quarterly. Figure 3.1 displays these series. I also use the US real and nominal GDP series (from the St. Louis Federal Reserve, BEA series A191RX1 and A191RC1) as q_t and y_t respectively, and construct $\Delta_t = q_t + \pi_q y_t$. The estimation requires setting two parameters, π_q and the time discount factor β. I set $\pi_q = 0.85$, the value in MW, and $\beta = 0.98$.

3.6.2 Estimation

I conduct the following maximum likelihood estimation in each regime. First, I model Δ_t as an ARMA process with intercept and linear trend, and use Akaike information criterion model selection to choose the order. This yielded an ARMA(1,1) specification. After transformation to the moving average representation, this provides estimates of parameters defining the Δ_t series: the MA coefficients $\{\alpha_{\Delta,l}\}$ and the residual variance σ^2_Δ. I then regress each p_{it} on Δ_t with an intercept and trend, with an ARMA(1,1) error structure. This yields the required p_{it} parameters, $\{\alpha_{z,l}\}$ and σ^2_z, and also the correlation between Δ_t and p_{it}. These models provide the ingredients for the identification result in Proposition 3.8. To obtain standard errors for the attention parameters, I use the delta method. I use the model specification given in Appendix B.8 that allows for negative partial correlations between Δ_t and p_{it}.

The attention parameters $\kappa_{\Delta,i}$, $\kappa_{z,i}$ and κ_i are in units of bits or nats. This is not an intuitive scale, and so instead I show results in terms of the ex post variance reduction. This is a quantity between 0 and 1 that gives the fraction of the prior variance that is explained by the signal:

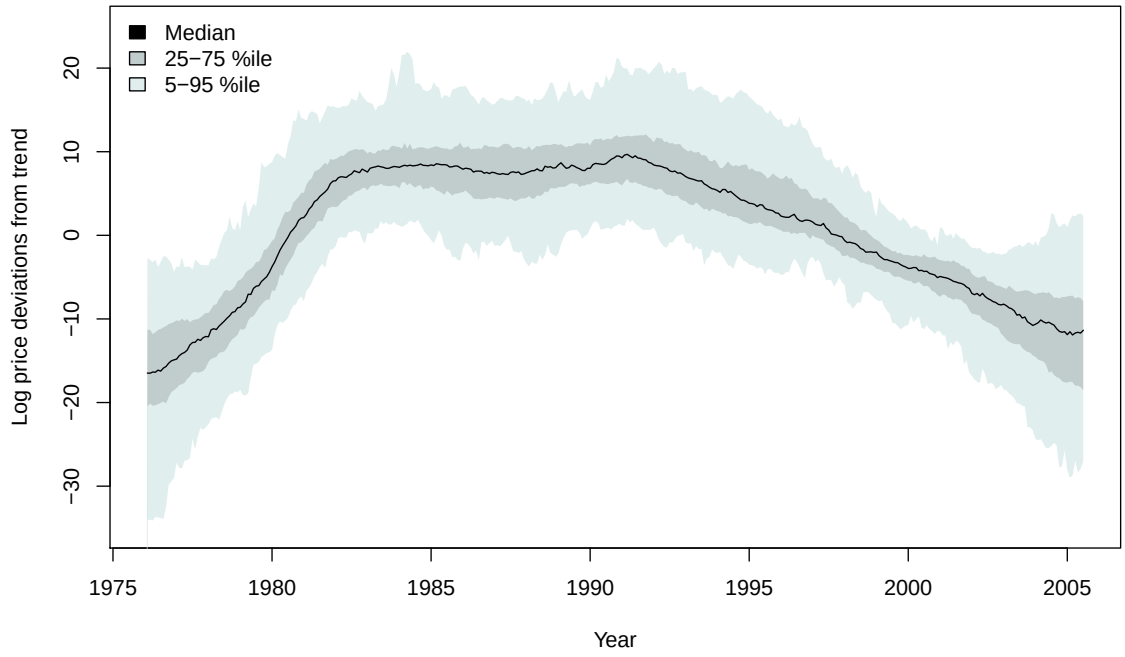


Figure 3.1: Disaggregated Personal Consumption Expenditure indices. These are 89 durable and nondurable goods sectors. Each is separately demeaned and linearly detrended. The data exhibit a break in linear trends: a pre-Volcker trend, a period of transition, and a Great Moderation trend.

$$VR_i \equiv 1 - \exp(-2\kappa_i) \tag{3.57}$$

$$VR_{\Delta,i} \equiv 1 - \exp(-2\kappa_{\Delta,i}) \tag{3.58}$$

$$VR_{z,i} \equiv 1 - \exp(-2\kappa_{z,i}) \tag{3.59}$$

For example, if $VR_i = 0$, then $\kappa_i = 0$ and the signal was completely uninformative. If $VR_i = 0.5$, then $\kappa_i = 0.35$ and 50% of the prior variance has been reduced by the signal. If $VR_i = 1$, then $\kappa = \infty$ and the signal was perfectly informative.

3.6.3 Regime definition

Figure 3.1 plots the disagreggated PCE data. The purpose of this display is simply to illustrate the general data shape across the whole sample: the decrease in volatility, possible change in MA coefficients, decrease in linear trend and shift in intercept. In

this plot only, each series is fit with a single linear trend and intercept for the entire sample. In the analysis that follows, I allow each series to have an independent trend and intercept in each regime. I fit these extra parameters as there is no reason to suppose that, when there is a regime change in MA coefficients and residual variance, that the trend parameters should remain constant. More generally, the RI model seeks to explain stochastic price movement around a medium- to long-run trend. The regimes that I fit are approximately a decade in length, and a trend for that length of time is longer than the kinds of short-run price movements that the MW model seeks to capture. Figure 3.1 suggests that there are multiple linear trends across the sample, and that estimates may be sensitive to regime specification.

I define the pre-Volcker regime as January 1976 to December 1983, consistent with the literature.[1] In defining the post-Volcker regime, it is necessary to allow a gap between regimes that is sufficiently large that agents have converged on steady-state behavior. The change in linear trends visible in Figure 3.1 suggests that a second regime definition should begin in 1990 or later. In Figure 3.2, I check the mean attention estimates for different definitions of the post-Volcker regime with a start date ranging from 1990 to 1995. In each regime I calculate the mean variance reduction, $\overline{VR} = \frac{1}{n}\sum_i VR_i$, and also $\overline{VR}_\Delta$ and $\overline{VR}_z$, defined analogously. In 1990 there is instability from the change in the linear trend, and in 1991 a year of adjustment. I selected the beginning of the second regime as January 1992, leaving a gap of 9 years between regimes. Using this regime definition, the detrended aggregate series in each regime are displayed in Figure 3.3.

3.6.4 ARMA estimates

Table 3.1 summarizes the maximum likelihood ARMA estimates for the series Δ_t and $\{p_{it}\}$. In Regime 2, the series display more persistence in the AR coefficient, but they have smaller residual variance. The mean linear trend in Regime 2 is

[1] I use the term 'pre-Volcker' loosely: although Volcker assumed Fed chairmanship in 1979, Kim and Nelson (1999) and McConnell and Perez-Quiros (2000) independently found a structural break in volatility in the first quarter of 1984.

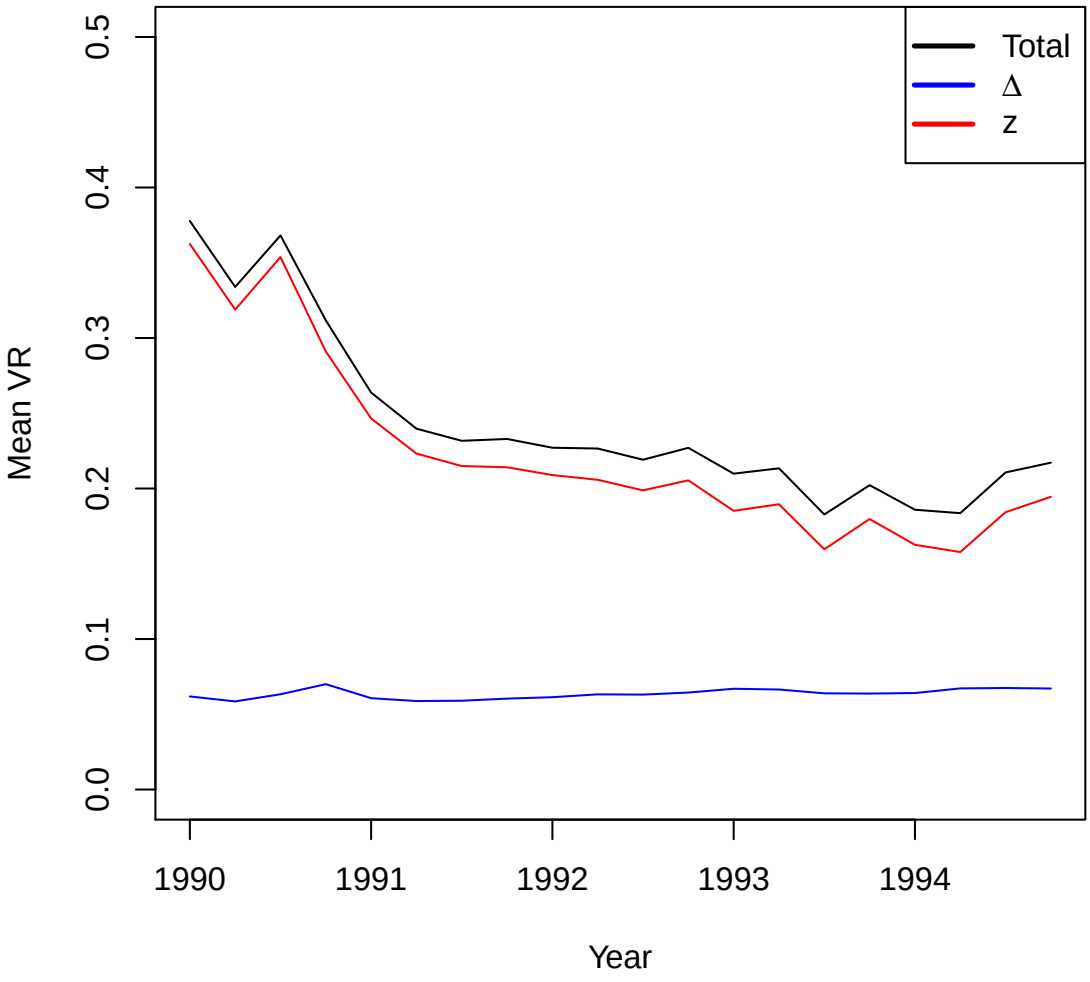


Figure 3.2: Mean ex post variance reduction, $\overline{VR} = \frac{1}{n}\sum_i VR_i$, and also $\overline{VR}_\Delta$ and $\overline{VR}_z$, for various definitions of Great Moderation regimes. For each time point in the plot, the regime with this start date and the end date of June 2005 was constructed. For each regime, the 3 mean variance reduction statistics were calculated. The purpose of the plot is to check whether agents have converged on steady state behavior at the start of the second regime. From this plot, I select a Regime 2 start period of January 1992, at which point the estimates have leveled.

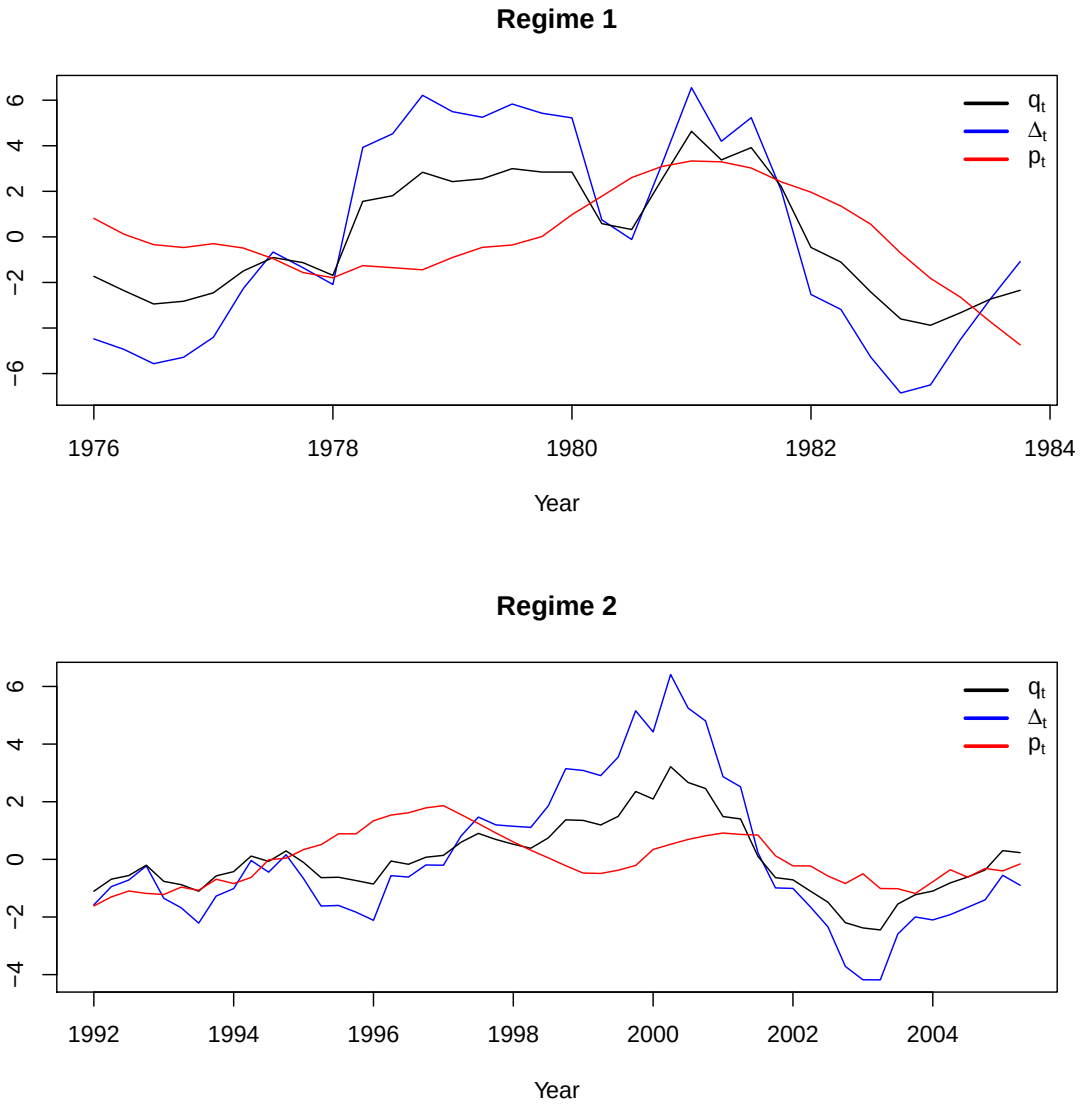


Figure 3.3: Aggregate series in each regime. The data are q_t (US GDP), y_t (US real GDP), and the 89 PCE price indices. The price level p_t is the mean of the PCE indices. Note that $\Delta_t = q_t + \pi_q y_t$.

lower, indicating the more moderate inflation level. The variation in linear trends among the sector price series is roughly similar in both regimes.

The estimation procedure relies on the assumption that agents pay attention to the current innovations of Δ_t. This in turn relies on the assumption that the coefficients in the moving average representation of Δ_t are nonincreasing. In Figure 3.4, I show the impulse response function in each regime given a standard deviation shock to q_t. Note that p_t is the mean of the 89 PCE indices. I compare the price

		AR coefficient	MA coefficient	Linear trend	Residual variance
Δ_t	Regime 1	0.808 (0.106)	0.289 (0.178)	0.029 (0.002)	0.00041
	Regime 2	0.924 (0.053)	0.100 (0.107)	0.020 (0.001)	0.00007
$\{p_{it}\}$	Regime 1	0.842 (0.157)	0.491 (0.281)	0.015 (0.005)	0.00031
	Regime 2	0.855 (0.150)	0.234 (0.225)	0.002 (0.006)	0.00028

Table 3.1: ARMA estimates for Δ_t and average ARMA parameters for the set of series $\{p_{it}\}$ by regime. Quantities in parentheses for Δ_t parameters are maximum likelihood standard errors. Quantities in parentheses for p_{it} parameters indicate the standard deviation of the point estimates.

level response predicted by the model with a response estimated from an SVAR with 10 lags identified by the causal ordering (q_t, Δ_t, p_t). There are a couple things to note. The first is that the assumption of nonincreasing MA coefficients on Δ is slightly violated; the response function peaks at period $t+1$. The lagged effect is only slightly higher than the impact effect, and one could argue that the estimation procedure provides a reasonable approximation. However, this violation would imply that in steady state, agents would pay attention to the first lagged innovation. Consequently, they do not devote their full attention capacity to the current innovation, and the estimate of attention capacity that is based on the contemporaneous covariance between sector prices and aggregate conditions would be biased downward. It is unclear how this bias would affect the low correlation in attention capacity that I find between regimes.

However, there is a more significant problem with the SVAR results. If the IRF of a shock in time t to q_t peaks in period $t+1$, then in steady state agents will only pay attention to the current and the first lagged innovation. Therefore, it is impossible under the model for the effect of this shock to peak after period $t+1$. However, in both regimes, the SVAR estimate of the response in p_t peaks with a delay, in period $t+7$ in Regime 1 and $t+5$ in Regime 2. This cannot be explained by the model as formulated.

3.6.5 Attention estimates

In Table 3.2, I show the estimates of mean variance reduction by regime. According to the model, the mean attention endowment increased from Regime 1 to Regime

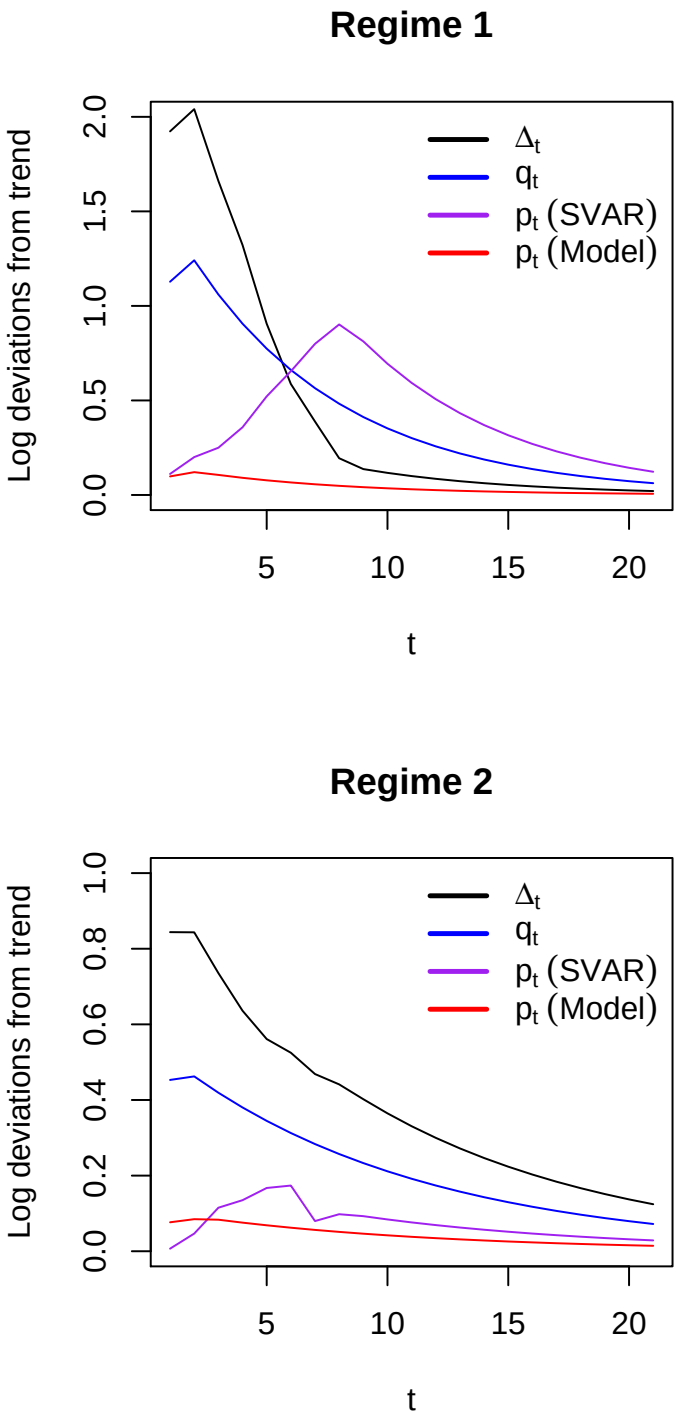


Figure 3.4: Impulse responses to a one a standard deviation shock to q_t. q_t and Δ_t are constructed from US nominal and real GDP. p_t is the mean of the 89 PCE indices. The model p_t response is produced from the MW model estimation. The SVAR p_t and the Δ_t response are produced from an SVAR with 10 lags identified by the causal ordering (q_t, Δ_t, p_t).

	$\overline{VR}_i$	$\overline{VR}_{\Delta,i}$	$\overline{VR}_{z,i}$
Regime 1	0.133 (0.253)	0.043 (0.106)	0.109 (0.231)
Regime 2	0.225 (0.326)	0.061 (0.182)	0.207 (0.328)

Table 3.2: Mean variance reduction estimates in each regime. Quantities in parentheses are sample standard deviations.

	Δ_t	$(\hat{\Delta}_{it} - \Delta_t)$	z_t	$(\hat{z}_{it} - z_t)$
Regime 1	0.019 (0.042)	0.542 (0.126)	0.051 (0.125)	0.388 (0.120)
Regime 2	0.015 (0.029)	0.552 (0.233)	0.133 (0.266)	0.300 (0.169)

Table 3.3: Variance decomposition by regime. The total variance in each price series p_{it} is decomposed into the contributions of Δ_t, z_t, and signal noises. The table displays the mean of these decompositions and, in parentheses, the standard deviation.

2. $\overline{VR}_{\Delta,i}$ increased by nearly half, but $\overline{VR}_{z,i}$ nearly doubled. Note that the misspecification discussed in the previous section will produce low attention estimates. The increase in estimates between regimes reflects a larger contemporaneous effect of Δ_t on p_{it} in Regime 2.

In Table 3.3, I show a summary of the variance decomposition of each series p_{it}. I estimate that Δ_t and the signal noise in $\hat{\Delta}_{it}$ account for a similar fraction of the mean variation in p_{it} in each regime. However, I estimate that z_{it} accounts for an increased fraction of the variation over time, and the signal noise in $\hat{z}_{it}$ accounts for a decreased fraction.

Figures 3.5 and 3.6 display the distribution of variance reduction point estimates and the accompanying delta method confidence intervals. Note that the parameters are weakly identified, and that a large fraction of attention parameters are estimated at levels close to 0.

Figure 3.7 displays the total variance reduction for each price series in Regime 1, with each series labeled. Food and fuel sectors, which are relatively volatile, tend to have larger attention estimates. Durable goods have smaller attention estimates. Figure 3.8 displays the difference in variance reduction between regimes, with each series labeled. The sectors with the largest increase in attention were largely household goods, and food and durable goods formed the bulk of the sectors with the largest decrease in attention estimates.

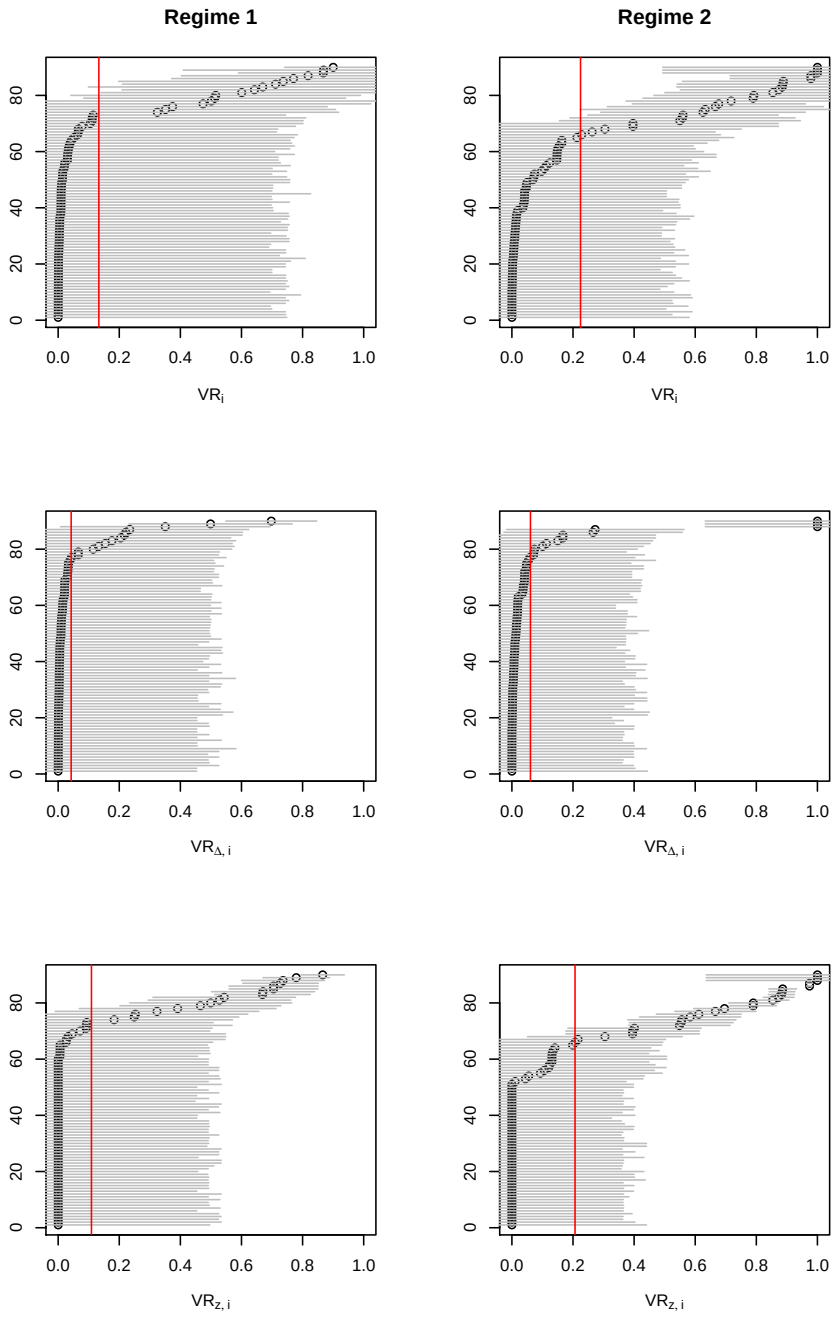


Figure 3.5: Point estimates of $\overline{VR}_i$, $\overline{VR}_{\Delta,i}$ and $\overline{VR}_{z,i}$, for each PCE series in both regimes. Grey bars indicate 95% confidence intervals. Vertical red lines indicate the mean point estimate.

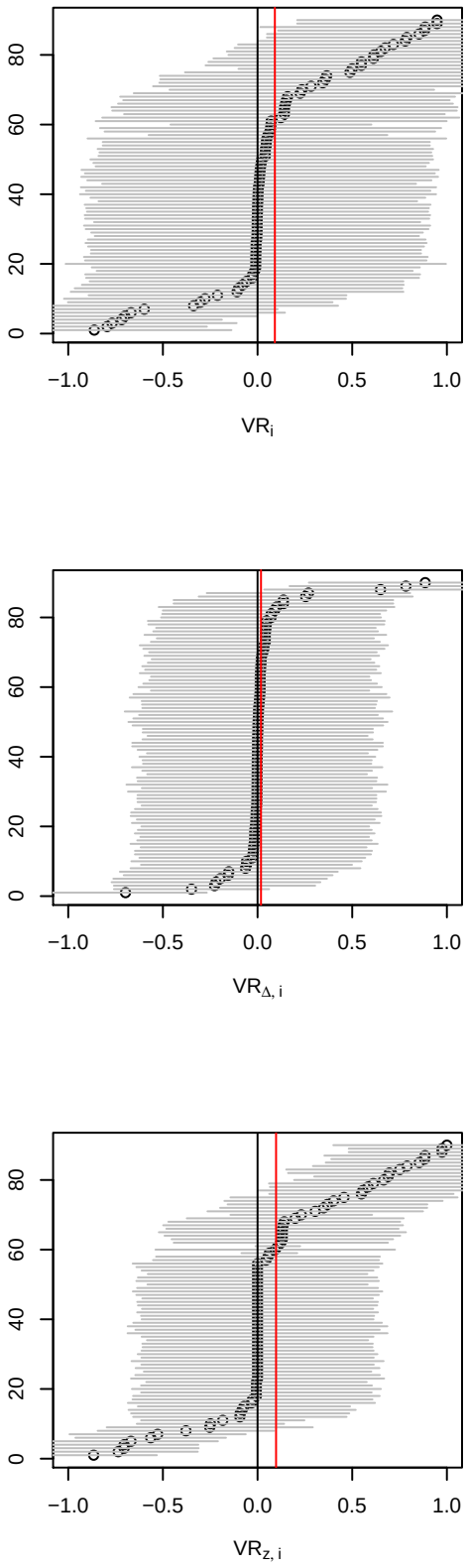


Figure 3.6: Point estimates for the change between regimes in each $\overline{VR}_i$, $\overline{VR}_{\Delta,i}$ and $\overline{VR}_{z,i}$. Grey bars indicate 95% confidence intervals. Vertical red lines indicate the mean point estimate.

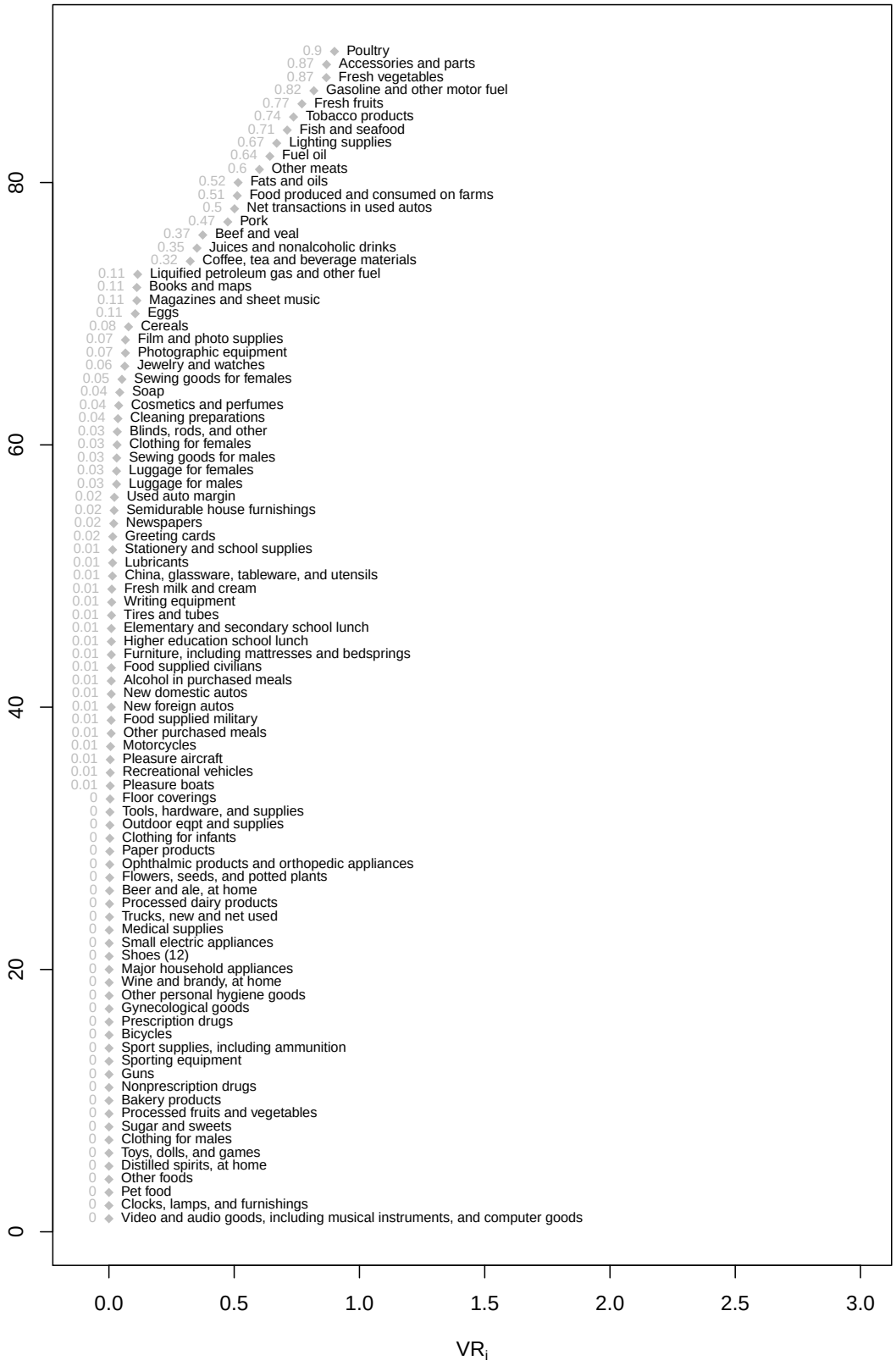


Figure 3.7: Point estimates of VR_i in Regime 1, with each PCE sector labeled.



Figure 3.8: The change in VR_i between regimes, with each PCE sector labeled.

Figure 3.9 shows the correlation between variance reduction for each price series between regimes. The correlation in these estimates between regimes is low, which may reflect the weak identification of these parameters, and may also reflect the model misspecification.

3.7 Conclusion

This chapter outlined an estimation procedure for a Rational Inattention model, and attempted to fit it to US price data. The goal of the estimation was to determine whether RI theory of attention reallocation provides predictive power, in the specific sense of the data producing attention capacities that are correlated between regimes. Although the attention capacity of a sector between regimes was not strongly correlated, there were a number of limitations of the modeling approach. The first was that the data violated the functional form assumption for Δ_t of nonincreasing moving average coefficients. The second was that the model did not explain the delayed reaction of the price level to an aggregate demand shock. This is likely due to the very simplified model, in which there is only one aggregate shock series. The third was that the estimation procedure only weakly identified the attention parameters, and produced large confidence intervals.

Given these caveats, there was some evidence that information capacity increased across regimes. It is plausible that the cost of gathering information, particularly of gathering idiosyncratic information, would have decreased across the sample. This analysis provides a caveat for calibration of rational information models that assume a constant capacity constraint.

Future work may be able to provide estimates that overcome the limitations of the present analysis. For instance, an analysis that does not require steady-state behavior could provide more reliable estimates. A richer model of aggregate dynamics could better capture price level dynamics. An alternative approach would be to directly infer the level of κ from microdata on information gathering costs. While this may only provide a rough guide to the level of κ as it exists in the

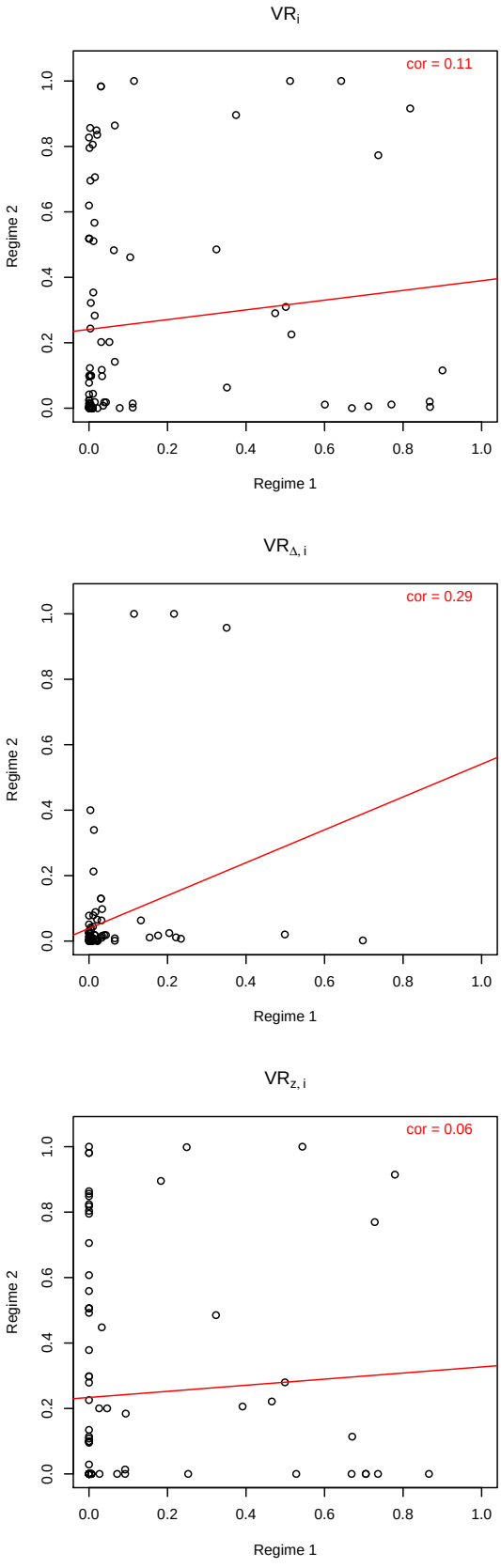


Figure 3.9: VR_i, $VR_{\Delta,i}$ and $VR_{z,i}$ in Regime 1 against the same estimates in Regime 2. The correlation between regimes and the linear trend are also displayed.

context of the MW model, it could at least provide evidence on whether changes in information capacity are plausible.

This paper is among the first to estimate an RI model. The key innovation of leveraging volatility regimes to test the optimality of agents' signal choices is original and could be a useful foundation for future work.

4

Social value of uncertainty

4.1 Introduction

There is empirical evidence that aggregate uncertainty increases in periods of high volatility (Bloom et al. 2012, Stock and Watson 2012), and this has led the news literature to investigate real effects of uncertainty shocks. A seminal paper in this literature is Bloom (2009), in which changes in the second moment of productivity shocks induce wait-and-see behavior among firms. Others have investigated a variety of different mechanisms. Pastor and Veronesi (2012) construct a general equilibrium model in which firms have uncertainty with respect to the impact of a government policy and also whether the policy regime will change in the future. Bloom et al. (2012) investigate uncertainty shocks as a source of business cycle fluctuations in a dynamic stochastic general equilibrium model. They note that higher uncertainty slows the pace that agents incorporate new information, meaning that policy interventions have delayed efficacy. Basu and Bundick (2017) model uncertainty shocks in a DSGE model with sticky prices around the zero lower bound. They find significant effects of uncertainty shocks via precautionary household saving.

The empirical work in this literature has used various uncertainty measures. Stock and Watson (2012) estimate a dynamic factor model and find evidence that uncertainty shocks played a principal role in the 2008 recession. Drechsler and

Yaron (2010) investigate how perceived market uncertainty is related to prices by constructing a measure of the variance premium and using it to predict stock returns. Baker et al. (2016) develop an index of economic policy uncertainty using newspaper article data. They show that this uncertainty spikes around significant news events (wars, government and economic crises) and predicts decreases in investment, output and employment.

In all of this work, uncertainty is a measure of the second moment. In the theoretical models listed above, agents know the current variance of shocks. In reality, however, there may be substantial uncertainty with respect to the current variance. This raises the question of how the inclusion of this kind of variance-uncertainty changes the behavior of a news model. In this vein, Beaudry and Portier (2014) suggest that uncertainty shocks could also be understood as uncertainty about shock variances, in the sense of expected squared future growth rates. To my knowledge, the work on this question is very limited. The working paper Berger et al. (2017) is the only study I am aware of that investigates the real effects of uncertain volatility. They take the definition of a news shock as the innovations in expected future TFP that are orthogonal to contemporaneous TFP innovations, and construct an analogous measure for the volatility of aggregate stock returns. They estimate a VAR to identify two variance series: realized volatility and the level of uncertainty about future stock returns. They define their second moment shock as the innovations in expected future stock returns that are orthogonal to contemporaneous volatility in stock prices. They find that realized volatility is correlated with output, consumption, investment, and employment. The uncertainty shock is not significantly correlated with any real variables, and the direction of association is not robust to different VAR specifications. They conclude that realized volatility does not affect real variables through uncertainty about the future. Either realized volatility affects real variables through a different channel, or there is a confounder, and another variable causes both volatility and changes to real variables simultaneously. They suggest a model with a negatively skewed shock distribution, which induces a correlation between the volatility and the level of the shock.

The purpose of this paper is to provide basic theoretical results about variance-uncertainty and how it can behave differently from mean-uncertainty. To clarify these terms, suppose that e follows a distribution with zero mean and stochastic variance σ_e^2. Further suppose that σ_e^2 follows a distribution with mean μ_σ and its variance τ_σ:

$$e \sim f_e(0, \sigma_e^2),\ \mathrm{E}(e) = 0,\ \mathrm{Var}(e) = \sigma_e^2$$
$$\sigma_e^2 \sim f_{\sigma_e^2}(\mu_\sigma, \tau_\sigma),\ \mathrm{E}(\sigma_e^2) = \mu_\sigma,\ \mathrm{Var}(\sigma_e^2) = \tau_\sigma$$

I call μ_σ mean-uncertainty and τ_σ variance-uncertainty.

I examine the behavior of both kinds of uncertainty in one of the simplest and most widely cited models in the news literature, Morris and Shin (2002). This model has been frequently cited in the monetary policy literature (Blinder et al., 2008), as well as the news literature, through its isomorphism to the Lucas islands model (for example, Nimark, 2014, Lorenzoni, 2009, and Woodford, 2003). For simplicity, the model is static, which is sufficient to illustrate the distinguishing features of the two kinds of uncertainty.

In Morris and Shin (2002) (henceforth MS), a continuum of agents report an estimate of a stochastic variable. Each agent receives a private signal and a public signal. The private signals have independent noise. The agents have a two-part objective, which contains a quadratic penalty in accuracy and also a zero-sum quadratic penalty in distance from the mean response (beauty-contest term). The relative weight placed on the beauty-contest term may vary. The beauty-contest term causes agents to overweight the public signal relative the welfare-optimizing strategy. I add the single modeling feature that agents do not know the variances of the public and the private signal with certainty, but have prior distributions over the variances. Each prior distribution is characterized by a mean and a variance, the mean- and variance-uncertainty. Varying the two kinds of uncertainty changes not only the precision of the agent's estimate but also the agents' overweighting of the public signal, impacting welfare.

I provide welfare results in the case where either the public or the private variance is known and fixed and the other is unknown and stochastic. By varying either the mean or the variance of the prior generating the stochastic variance, one can investigate the effects of mean- and variance-uncertainty. The results for mean-uncertainty are similar to the MS results. Welfare is decreasing in the mean-uncertainty of the private signal. However, welfare may be decreasing or increasing in the mean-uncertainty of the public signal, depending on parameter values. The effect of variance-uncertainty is, in a sense, the opposite. Welfare is increasing in the variance-uncertainty of the private signal. However, welfare may be decreasing or increasing in the variance-uncertainty of the public signal, depending on parameter values.

Section 4.2 outlines the model, Section 4.3 gives the welfare results, Section 4.4 illustrates the behavior of the model, and Section 4.5 concludes by discussing this model in the context of the literature on central bank transparency.

4.2 Model

In this section, I first outline the MS model (Section 4.2.1), and then I outline the generalized model with unknown variances (Section 4.2.2).

4.2.1 MS model

This section directly follows MS. Suppose there is a continuum of agents indexed by the unit inverval $[0, 1]$, and there is an exogenous stochastic variable Θ with realization θ and an improper uniform prior over $\mathbb{R}$. Agent $i \in [0, 1]$ chooses an action $a_i \in \mathbb{R}$. Let $\mathbf{a}$ denote the action profile over all agents, and let the mean action be denoted $\bar{a} = \int_0^1 a_i di$. Agent i has the utility function

$$u_i(\mathbf{a}, \theta) \equiv -(1 - r)(a_i - \theta)^2 - r(L_i - \bar{L}) \tag{4.1}$$

where r is a constant, such that $0 \leq r < 1$, and

$$L_i \equiv \int_0^1 (a_j - a_i)^2 dj \tag{4.2}$$

$$\bar{L} \equiv \int_0^1 L_j dj \tag{4.3}$$

The first term in (4.1) is the accuracy term, a quadratic loss in deviation from θ. The second term in (4.1) is quadratic loss in deviation from the mean action $\bar{a}$, the 'beauty-contest' term. It is recentered so that it sums to zero across all agents. This third recentering term is clearer in the following expansion, plugging (4.2) into (4.1):

$$u_i(\mathbf{a}, \theta) = -(1-r)(a_i - \theta)^2 - r\left(\int_0^1 (a_j - a_i)^2 dj - \int_0^1 \int_0^1 (a_k - a_j)^2 dk dj\right) \tag{4.4}$$

The first term is the accuracy penalty, the second term is the beauty-contest penalty, and the third term is a recentering term that does not depend on a_i, such that the beauty-contest penalties sum to 0 across agents. Let $\mathcal{I}_i$ denote the agent's information set, and let $\mathrm{E}_i(\cdot) = \mathrm{E}(\cdot|\mathcal{I}_i)$ denote the agent's expectation conditional on their information set. The agent maximizes expected welfare: $\max_{a_i}(\mathrm{E}_i(u_i(\mathbf{a}, \theta)))$. The first order condition of this problem is a weighted average between the agent's estimate of θ and the agent's estimate of the mean action $\bar{a}$:

$$a_i = (1-r)E_i(\theta) + rE_i(\bar{a}) \tag{4.5}$$

The welfare function is the negative mean squared error of the agents' actions, which can be written as the scaled mean of agent utilities:

$$W(\mathbf{a}, \theta) \equiv \frac{1}{1-r} \int_0^1 u_i(\mathbf{a}, \theta) di \tag{4.6}$$

$$= -\int_0^1 (a_i - \theta)^2 di \tag{4.7}$$

Agents each receive two signals. All agents receive a common public signal with zero-mean Gaussian noise η that has a precision α:

$$y = \theta + \eta, \eta \sim N(0, \alpha^{-1}) \tag{4.8}$$

Each agent also receives a private signal with zero-mean Gaussian noise ϵ_i that has a precision β:

$$x_i = \theta + \epsilon_i, \epsilon_i \sim N(0, \beta^{-1}) \tag{4.9}$$

The private errors are pairwise independent such that $\mathrm{E}(\epsilon_i \epsilon_j) = 0,\ \forall i \neq j$. Note that the welfare-optimizing response is the agent's conditional expectation $\mathrm{E}_i(\theta)$, which is the precision-weighted average of y and x_i:

$$\mathrm{E}_i(\theta) = \frac{\alpha y + \beta x_i}{\alpha + \beta} \tag{4.10}$$

However, unless agents place zero weight on the beauty-contest term ($r = 0$), or unless either the private or the public signal is noiseless ($\alpha = \infty$ or $\beta = \infty$), there will be a coordination failure. Agents will anticipate that other agents will place weight on the public signal y, and consequently overweight y in their response a_i. MS show that the unique equilibrium action a_i of agent i satisfies

$$a_i = \frac{\alpha y + (1-r)\beta x_i}{\alpha + (1-r)\beta} \tag{4.11}$$

Expected welfare, denoted $\mathrm{E}[W(\mathbf{a}, \theta)]$, is equal to

$$\mathrm{E}[W(\mathbf{a}, \theta)] = -\frac{\alpha + \beta(1-r)^2}{[\alpha + \beta(1-r)]^2} \tag{4.12}$$

For proof of uniqueness and existence of the equilibrium, see MS.

The MS welfare results focus on the regions of the parameter space (α, β) in which expected welfare is increasing. We can examine the partial derivatives of expected welfare with respect to each precision parameter:

$$\frac{\partial \mathrm{E}[W(\mathbf{a}, \theta)]}{\partial \alpha} = \frac{\alpha - (2r-1)(1-r)\beta}{[\alpha + \beta(1-r)^3]} \geq 0 \Leftrightarrow \frac{\alpha}{\beta} \leq \frac{1}{(2r-1)(1-r)} \tag{4.13}$$

$$\frac{\partial \mathrm{E}[W(\mathbf{a}, \theta)]}{\partial \beta} = \frac{(1-r)[(1+r)\alpha + (1-r)^2\beta]}{[\alpha + \beta(1-r)]^3} > 0,\ \forall\ (\alpha, \beta) \in \mathbb{R}^2_+ \tag{4.14}$$

The expected welfare is always increasing in the precision of the private signal, β. However, in cases where the public signal is imprecise relative to the private

signal and the weight on the beauty contest term is high, then it is possible for the expected welfare to be decreasing in the precision of the public signal.

4.2.2 Generalized model

Model specification

Suppose the model is the same as in (4.1) - (4.9), except that α or β are stochastic and unknown to the agent. Let α, β, denote realizations of the random variables A, B. Agents have a common prior over these distributions: $A \sim f_A(\alpha),\ B \sim f_B(\beta)$.

The MS model above is a special case of this generalized model: if either α or β (or both) are known to the agent, $f_A(\alpha)$ or $f_B(\beta)$ is a Dirac delta distribution with point mass at α or β.

As before, the agent receives the signals $\{x_i, y\}$ as specified in (4.8) and (4.9). Given these signals, the agent calculates a posterior distribution $f_{\Theta,A,B|\mathcal{I}_i}(\theta, \alpha, \beta)$. Let $\phi(\cdot|\mu, \sigma^2)$ denote a Gaussian density with mean μ and variance σ^2. The agent's posterior is the scaled product of the likelihood and the priors:

$$f_{\Theta,A,B|\mathcal{I}_i}(\theta, \alpha, \beta) = \frac{\phi(y|\theta, \alpha^{-1})\phi(x_i|\theta, \beta^{-1})f_A(\alpha)f_B(\beta)}{f(\mathcal{I}_i)} \tag{4.15}$$

I show in Appendix C.1 that the normalizing constant $f(\mathcal{I}_i)$ is:

$$f(\mathcal{I}_i) = \int_0^\infty \int_0^\infty \phi(y - x_i|0, \alpha^{-1} + \beta^{-1})f_A(\alpha)f_B(\beta)d\alpha d\beta \tag{4.16}$$

The agent's marginal posterior for θ and (α, β) are:

$$f_{\Theta|\mathcal{I}_i}(\theta) = \int_0^\infty \int_0^\infty f_{\Theta,A,B|\mathcal{I}_i}(\theta, \alpha, \beta)d\alpha d\beta \tag{4.17}$$

$$f_{A,B|\mathcal{I}_i}(\alpha, \beta) = \int_{-\infty}^\infty f_{\Theta,A,B|\mathcal{I}_i}(\theta, \alpha, \beta)d\theta \tag{4.18}$$

The posterior expectation of θ is:

$$E_i(\theta) \equiv \int_{-\infty}^\infty \theta f_{\Theta|\mathcal{I}_i}(\theta)d\theta \tag{4.19}$$

Each agent i also forms a posterior density for another agent's signal $x_j,\ i \neq j$:

$$f_{X_j|\mathcal{I}_i}(x_j) = \int_0^\infty \int_0^\infty \int_{-\infty}^\infty \phi(x_j; \theta, \beta^{-1}) f_{\Theta,A,B|\mathcal{I}_i}(\theta, \alpha, \beta) d\theta d\alpha d\beta \tag{4.20}$$

These quantities provide the ingredients to solve for the equilibrium action schedule $\mathbf{a}$.

It is still the case that the agent's first order condition is as in (4.5):

$$a_i = (1-r)E_i(\theta) + rE_i(\bar{a}) \tag{4.21}$$

Expanding this condition, where $\mathrm{E}_i(\theta)$ and $f_{X_j|\mathcal{I}_i}(x_j)$ are as defined in (4.19) and (4.20):

$$a_i = r\mathrm{E}_i(\theta) + (1-r)\int_{-\infty}^\infty a_j f_{X_j|\mathcal{I}_i}(x_j) dx_j \tag{4.22}$$

This is an inhomogenous Fredholm equation of the second kind. The action profile $\mathbf{a}$ defined by this condition does not generally have an analytic solution. In Appendix C.2, I prove that there exists a function a_i in the space of continuous functions of $\mathcal{I}_i$ that satisfies (4.22). This solution is not necessarily unique.

Prior specification

In the welfare results, I consider the special case of the model in which only one of either A^{-1} or B^{-1} is stochastic, and the other precision is fixed. When A^{-1} is stochastic, I assign it an inverse Gamma prior distribution (equivalently, assign A a Gamma prior), and the same for B^{-1}. The inverse Gamma distribution is the standard choice for variance parameters with support on $\mathbb{R}^+$ (Harrison and West, 1999).

Below is a specification for $f_{A^{-1}}(\alpha^{-1})$, and $f_{B^{-1}}(\beta^{-1})$ can be defined analogously.

Let the variance A^{-1} follow an inverse Gamma distribution with shape and scale parameters $k_\alpha > 2, \ell_\alpha > 0$:

$$f_{A^{-1}}(\alpha^{-1}) = \frac{\ell_\alpha^{k_\alpha}}{\Gamma(k_\alpha)} (\alpha^{-1})^{-k_\alpha - 1} \exp\left(-\frac{\ell_\alpha}{\alpha^{-1}}\right) \tag{4.23}$$

The mean and the variance of of α^{-1} are defined as follows, with support $\mu_\alpha, \tau_\alpha \in \mathbb{R}^+$:

$$\mu_\alpha = \mathrm{E}(\alpha^{-1}) = \frac{\ell_\alpha}{k_\alpha - 1} \tag{4.24}$$

$$\tau_\alpha = \mathrm{Var}(\alpha^{-1}) = \frac{\ell_\alpha^2}{(k_\alpha - 1)^2(k_\alpha - 2)} \tag{4.25}$$

Equivalently,

$$k_\alpha = \frac{\mu_\alpha^2 + 2\tau_\alpha}{\tau_\alpha},\ \ell_\alpha = \frac{\mu_\alpha(\mu_\alpha^2 + \tau_\alpha)}{\tau_\alpha} \tag{4.26}$$

Figure 4.1 plots this distribution for a range of paremeters. Although I illustrate the model with a specific choice of prior, the welfare results only require the following properties of $f_{A^{-1}}(\alpha^{-1})$:

$$\exists\ \frac{\partial f_{A^{-1}}(\alpha^{-1})}{\partial \mu_\alpha},\ \mu_\alpha \in \mathbb{R}^+, \alpha \in \mathbb{R}^+ \tag{4.27}$$

$$\exists\ \frac{\partial f_{A^{-1}}(\alpha^{-1})}{\partial \tau_\alpha},\ \tau_\alpha \in \mathbb{R}^+, \alpha \in \mathbb{R}^+ \tag{4.28}$$

$$\frac{\partial F_{A^{-1}}(\alpha^{-1})}{\partial \mu_\alpha} < 0,\ \mu_\alpha \in \mathbb{R}^+, \alpha \in \mathbb{R}^+ \tag{4.29}$$

$$\tau_\alpha < \tau'_\alpha \Rightarrow f_{A^{-1}|\mu_\alpha,\tau'_\alpha}(\alpha^{-1}) \text{ is a mean-preserving spread of } f_{A^{-1}|\mu_\alpha,\tau_\alpha}(\alpha^{-1}) \tag{4.30}$$

(4.27) and (4.28) require that the prior is differentiable by its mean and variance. (4.29) states that one prior distribution will stochastically dominate another prior with a lower mean and the same variance. (4.30) states that changing the prior variance τ_α does not change the prior mean μ_α.

In addition to these properties, I assume that a_i as a function of x_i is smooth in the following sense: (1) a_i is differentiable in μ_α, μ_β, τ_α; (2) the cross derivative of a_i in r and τ_α exists; and (3) the cross derivative of a_i in r and τ_β exists. These assumptions are restated in Appendix C.3.1.

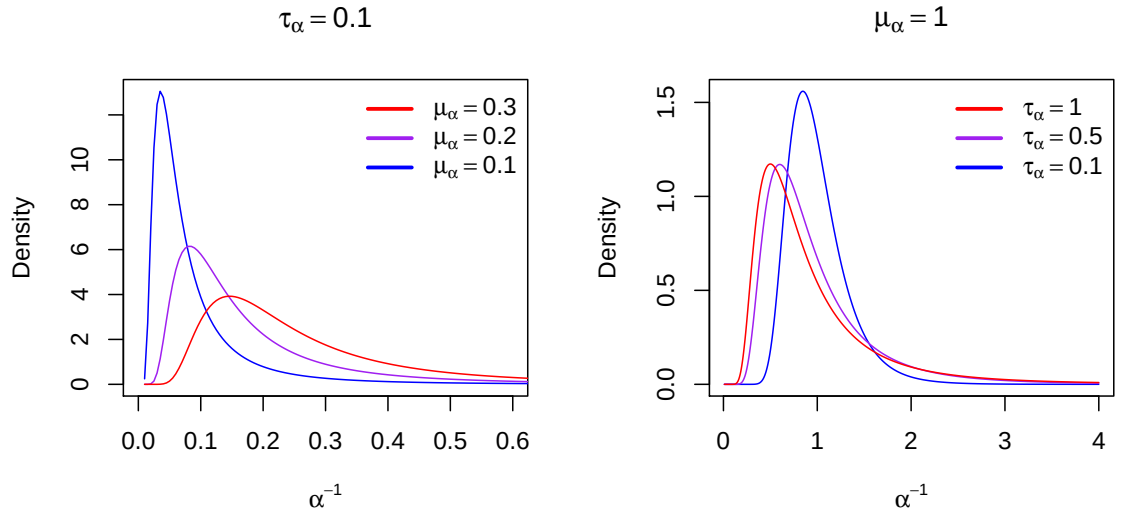


Figure 4.1: The prior distribution $f_{A^{-1}}(\alpha^{-1})$ for different values of μ_α and τ_α.

4.3 Welfare results

In this section, I outline the welfare results of the paper. The results are summarized using a decomposition of the expected welfare:

$$\begin{aligned} \mathrm{E}[W(a,\theta)] &= -\mathrm{E}[(a_i-\theta)^2] && (4.31)\\ &= -\mathrm{E}[((a_i-\mathrm{E}_i(\theta))+(\mathrm{E}_i(\theta)-\theta))^2] && (4.32)\\ &= -\mathrm{E}[(a_i-\mathrm{E}_i(\theta))^2]-\mathrm{E}[(\mathrm{E}_i(\theta)-\theta)^2] && (4.33) \end{aligned}$$

This follows from the fact that the estimation error $(\mathrm{E}_i(\theta)-\theta)$ is uncorrelated with functions of $\mathcal{I}_i$.

Now differentiate with respect to some parameter $(\cdot)$:

$$\frac{\partial \mathrm{E}[W(a,\theta)]}{\partial(\cdot)} = -\frac{\partial \mathrm{E}[(a_i-\mathrm{E}_i(\theta))^2]}{\partial(\cdot)} - \frac{\partial \mathrm{E}[(\mathrm{E}_i(\theta)-\theta)^2]}{\partial(\cdot)} \qquad (4.34)$$

The welfare effect of a parameter depends on two parts: the effect on the distortion in a_i induced by the beauty-contest term, and also the effect on the precision of $\mathrm{E}_i(\theta)$. I show the effects of the parameters $\{\mu_\alpha, \mu_\beta, \tau_\alpha, \tau_\beta\}$ on these two welfare terms. The results are summarized in Table 4.1.

	Fixed β		Fixed α	
	$\partial(\cdot)/\partial\mu_\alpha$	$\partial(\cdot)/\partial\tau_\alpha$	$\partial(\cdot)/\partial\mu_\beta$	$\partial(\cdot)/\partial\tau_\beta$
$-\partial\mathrm{E}[(a_i-\theta)^2]/\partial(\cdot)$	$+/-$	$+/-$	$-$	$+$
$-\partial\mathrm{E}[(a_i-\mathrm{E}_i(\theta))^2]/\partial(\cdot)$	$+/-$	$+/-$	$-$	$+$
$-\partial\mathrm{E}[(\mathrm{E}_i(\theta)-\theta)^2]/\partial(\cdot)$	$-$	$+$	$-$	$+$

Table 4.1: The effect of mean- and variance-uncertainty on the welfare terms. The welfare terms are rows (overall welfare, beauty-contest distortion, and estimation error). The prior parameters are in the columns. The + and − sign corresponds to a positive or negative gradient. The entry $+/-$ indicates that either a positive or a negative derivative is possible, depending on the parameter combination.

Note that x_i and y enter symmetrically into the function $\mathrm{E}_i(\theta)$; consequently, the third row of the table is the same for α and β. This row indicates that increasing the mean-uncertainty (higher μ_α or μ_β) always reduces the precision of $\mathrm{E}_i(\theta)$ (lower $-\mathrm{E}[(\mathrm{E}_i(\theta)-\theta)^2]$).

Increasing the variance-uncertainty (τ_α or τ_β) increases the precision of $\mathrm{E}_i(\theta)$ (higher $-\mathrm{E}[(\mathrm{E}_i(\theta)-\theta)^2]$). For an intuition about why this is true, consider the estimator that agents use when $\tau_\alpha=\tau_\beta=0$. The estimator is just a precision-weighted average: $\mathrm{E}_i^0(\theta)=(\mu_\alpha^{-1}y+\mu_\beta^{-1}x_i)/(\mu_\alpha^{-1}+\mu_\beta^{-1})$. When $\tau_\alpha=\tau_\beta=0$, then this estimator has the mean squared error

$$\mathrm{E}[(\mathrm{E}_i^0(\theta)-\theta)^2]=\mathrm{E}\left[\left(\frac{\mu_\alpha^{-1}y+\mu_\beta^{-1}x_i}{\mu_\alpha^{-1}+\mu_\beta^{-1}}-\theta\right)^2\right] \tag{4.35}$$

$$=\mathrm{E}\left[\left(\frac{\mu_\alpha^{-1}\eta+\mu_\beta^{-1}\epsilon_i}{\mu_\alpha^{-1}+\mu_\beta^{-1}}\right)^2\right] \tag{4.36}$$

$$=\left(\frac{\mu_\alpha^{-1}}{\mu_\alpha^{-1}+\mu_\beta^{-1}}\right)^2\mathrm{Var}(\eta)+\left(\frac{\mu_\beta^{-1}}{\mu_\alpha^{-1}+\mu_\beta^{-1}}\right)^2\mathrm{Var}(\epsilon_i) \tag{4.37}$$

$$=\frac{1}{\mu_\alpha^{-1}+\mu_\beta^{-1}} \tag{4.38}$$

Now consider the case where either $\tau_\alpha\neq 0$ or $\tau_\beta\neq 0$. It will still be the case that, if agents used the naive estimator $\mathrm{E}_i^0(\theta)$, then (4.35) - (4.38) would hold. This is because $\mathrm{Var}(\eta)=\mu_\alpha$ and $\mathrm{Var}(\epsilon_i)=\mu_\beta$ are unchanged. But the naive estimator is no longer optimal. It has failed to incorporate the information about the variances α^{-1} and β^{-1} that is contained in $\mathcal{I}_i$. As the variance-uncertainty

τ_α and τ_β increase, there is increasingly more information in the signals $\{y, x_i\}$ about α^{-1} and β^{-1}. Increasing the variance-uncertainty (τ_α or τ_β) increases the precision of $\mathrm{E}_i(\theta)$ (higher $-\mathrm{E}[(\mathrm{E}_i(\theta) - \theta)^2]$).

The effects of α and β are different for the beauty-contest distortion, the first row of Table 4.1. Increasing the private signal mean-uncertainty μ_β worsens the distortion (lower $-\mathrm{E}[(a_i - \mathrm{E}_i(\theta))^2]$). This is the same result as is found in MS. The intuition is that, when μ_β is higher, then agents place less weight on their private signal x_i in forming their estimate $\mathrm{E}_i(\theta)$. They also know that all other agents will also place less weight on their respective private signals. As a result, the welfare loss due to beauty-contest distortion increases.

The variance-uncertainty of the private signal τ_β acts in the opposite way on beauty-contest distortion. When τ_β is large, for x_i close to y, then agents believe that β is high, i.e. x_i is precise, and give x_i more weight in their response a_i. This effect is stronger for larger τ_β (when the prior on β is less informative). When agents calculate their higher-order beliefs about other agents' behavior, they know that, the closer another agent's x_i is to y, the greater the weight the agent will place on their x_i. The cumulative effect of this is to decrease the overweighting of y in the action schedule, and reduce the beauty-contest distortion.

The effects of public signal mean- and variance-uncertainty, μ_α and τ_α, are ambiguous. This is because these parameters have ambiguous beauty-contest distortion. For some parameterizations, the effect on beauty-contest distortion behaves with the same logic as described for μ_β and τ_β above, with the opposite effect sign. In these cases, increasing μ_α reduces beauty-contest distortion, and τ_α worsens beauty-contest distortion. However, it is possible for the effect to go in the opposite direction, as is described in MS, as in (4.13).

The following propositions restate these results.

Proposition 4.1 ***Welfare effect of*** μ_α**.** *Suppose the model is defined as in (4.15) - (4.26) and set* $\tau_\beta = 0$ *(β is fixed). Assuming the differentiability conditions for* a_i *given in Appendix (C.3.1), then*

$$\frac{\partial E[(E_i(\theta) - \theta)^2]}{\partial \mu_\alpha} > 0 \tag{4.39}$$

Proof: See Lemma C.7 in Appendix C.4. ∎

Proposition 4.2 ***Welfare effect of*** μ_β**.** *Suppose the model is defined as in (4.15) - (4.26) and set* $\tau_\alpha = 0$ *(*α *is fixed). Assuming the differentiability conditions for* a_i *given in Appendix (C.3.1), then*

$$\frac{\partial E[(a_i - \theta)^2]}{\partial \mu_\beta} \geq 0 \text{ and } \frac{\partial E[(E_i(\theta) - \theta)^2]}{\partial \mu_\beta} \geq 0 \tag{4.40}$$

Proof: See Lemmas C.9 and C.10 in Appendix C.4. ∎

Proposition 4.3 ***Welfare effect of*** τ_α**.** *Suppose the model is defined as in (4.15) - (4.26) and set* $\tau_\beta = 0$ *(*β *is fixed). Assuming the differentiability conditions for* a_i *given in Appendix (C.3.1), then*

$$\frac{\partial E[(E_i(\theta) - \theta)^2]}{\partial \tau_\alpha} \leq 0 \tag{4.41}$$

Proof: See Lemma C.11 in Appendix C.4. ∎

Proposition 4.4 ***Welfare effect of*** τ_β**.** *Suppose the model is defined as in (4.15) - (4.26) and set* $\tau_\alpha = 0$ *(*α *is fixed). Assuming the differentiability conditions for* a_i *given in Appendix (C.3.1), then*

$$\frac{\partial E[(a_i - \theta)^2]}{\partial \tau_\beta} \leq 0 \text{ and } \frac{\partial E[(E_i(\theta) - \theta)^2]}{\partial \tau_\beta} \leq 0 \tag{4.42}$$

Proof: See Lemmas C.12 and C.14 in Appendix C.4. ∎

4.4 Example

In this section, I illustrate the behavior of the model. The model has no location in the sense that θ is uniformly distributed over $\mathbb{R}$, so for display purposes I normalize the location by fixing $y = 0$. I normalize scale with $\mu_\beta = 1$ in the cases where β is stochastic, and $\beta = 1$ when it is fixed. As benchmark parameters, I consider levels of μ_α (when α is stochastic) or α^{-1} (when α is fixed) of 0.3, 0.2, and 0.1.[1] I vary τ from 0 (no variance-uncertainty) to 1, a level which I found sufficiently large that further increases in τ had negligible effects on model behavior. I compare two levels of r, 0.25 and 0.75.

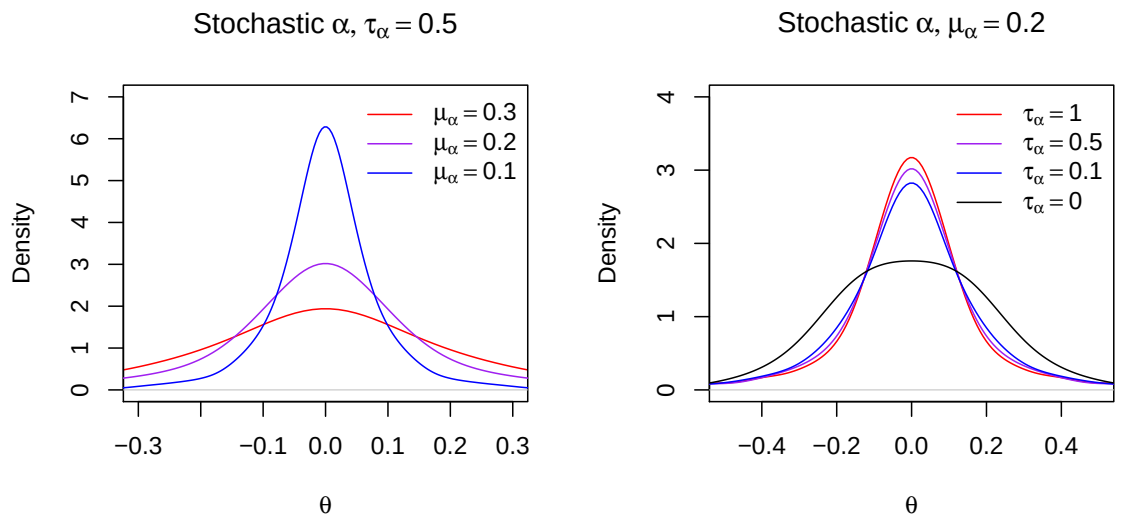


Figure 4.2: The distribution of Θ under different values of μ_α and τ_α.

Figure 4.2 shows the marginal distribution of θ when α is stochastic, for various levels of μ_α and τ_α. In the left-hand plot τ_α is held constant and μ_α is varied, and in the right-hand plot μ_α is held constant and τ_α is varied. This illustrates the effect of either changing the second moment versus the fourth moment of the distribution. Varying mean-uncertainty and varying variance-uncertainty have,

[1] These levels were chosen using a benchmark taken from Nimark (2014). Using his notation, he estimates a public signal variance of $\pi\sigma_\eta^2 = 0.051 * 0.49^2 = 0.012$ and a private signal variance of $\sigma_\epsilon^2 = 0.25^2 = 0.065$, which is a public-to-private noise variance ratio of 0.185.

in a sense, opposite effects. Decreasing the mean-uncertainty or increasing the variance-uncertainty makes the distribution of θ more peaked.

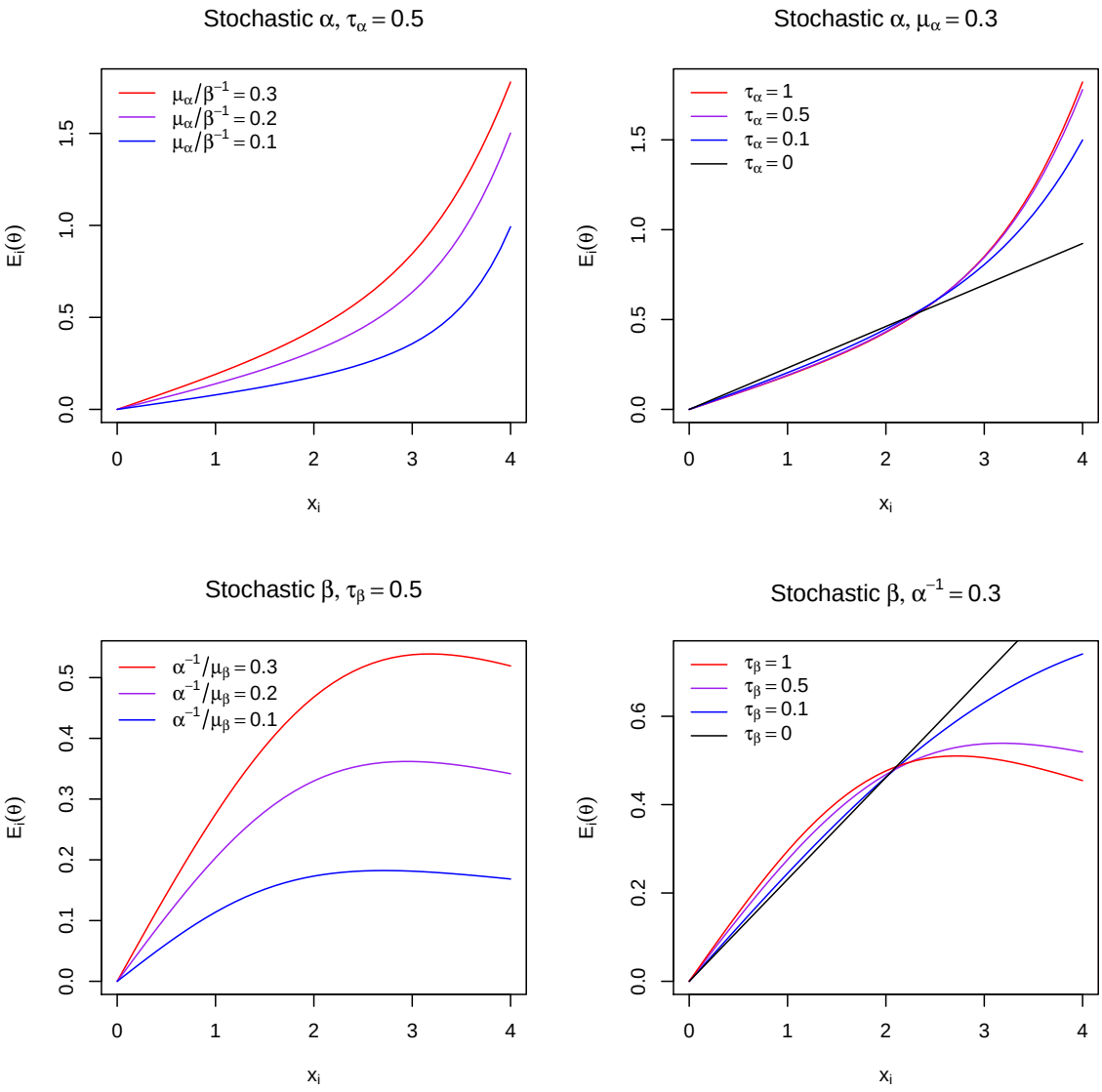


Figure 4.3: $\mathrm{E}_i(\theta)$ as a function of x_i, under different values of μ_α, τ_α, μ_β, and τ_β.

Figure 4.3 shows $\mathrm{E}_i(\theta)$ as a function of x_i under different parameter combinations (these plots only show positive x_i for clarity, as the function is symmetric about the origin). Because β and μ_β are normalized to 1, in the upper-left plot μ_α is varied, and in the lower-left plot α^{-1} is varied. Increasing the relative expected precision of the private signal increases the weight on the private signal (upper-left and lower-left). Increasing the variance-uncertainty of the public signal causes

$\mathrm{E}_i(\theta)$ to have greater convexity in $x_i > 0$ (and concavity in $x_i < 0$). Conversely, increasing the variance-uncertainty of the private signal causes $\mathrm{E}_i(\theta)$ to have greater concavity in $x_i > 0$ (and convexity in $x_i < 0$). This illustrates that, when there is more variance-uncertainty, agents' estimates of the stochastic precision are more sensitive to the realization of x_i.

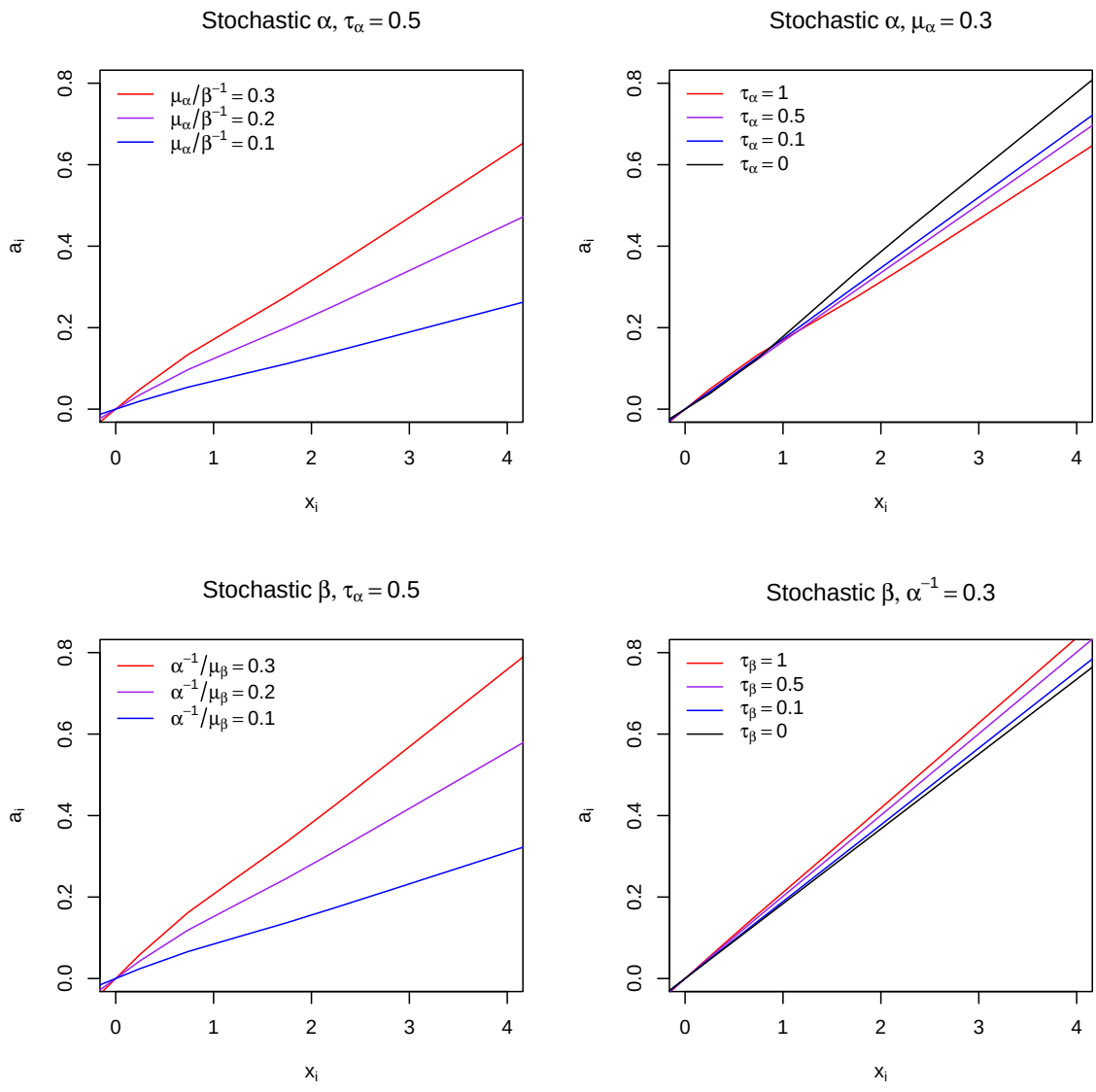


Figure 4.4: The action schedules for $r = 0.25$ under different values of μ_α, τ_α, μ_β, and τ_β.

Figures 4.4 and 4.5 show the action schedule under different parameter combinations (again, only positive x_i are shown). I estimate the solution to (4.22) with the

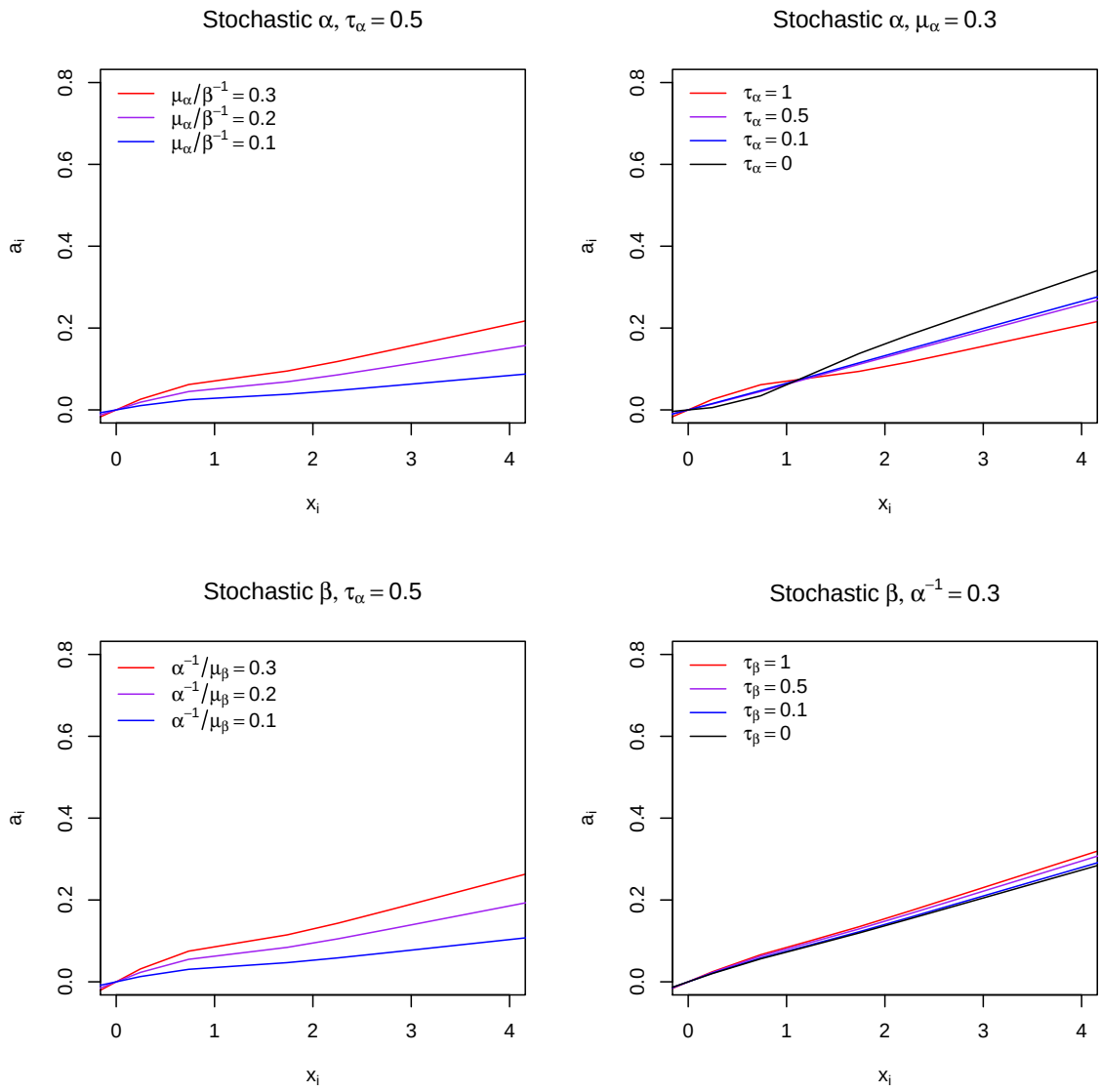


Figure 4.5: The action schedules for $r = 0.75$ under different values of μ_α, τ_α, μ_β, and τ_β.

Nyholm method, using Hermite weights (for estimation details, see Sloan (1981)). As with $\mathrm{E}_i(\theta)$, increasing the relative expected precision of the private signal increases the weight on the private signal (upper- and lower-left). For this parameterization, increasing the variance-uncertainty of the public signal decreases the magnitude of a_i (upper-right). Increasing the weight of the private signal increases the magnitude of a_i (lower-right). Lastly, increasing r decreases the magnitude of a_i.

Figures 4.6 and 4.7 show the expected welfare under different parameter

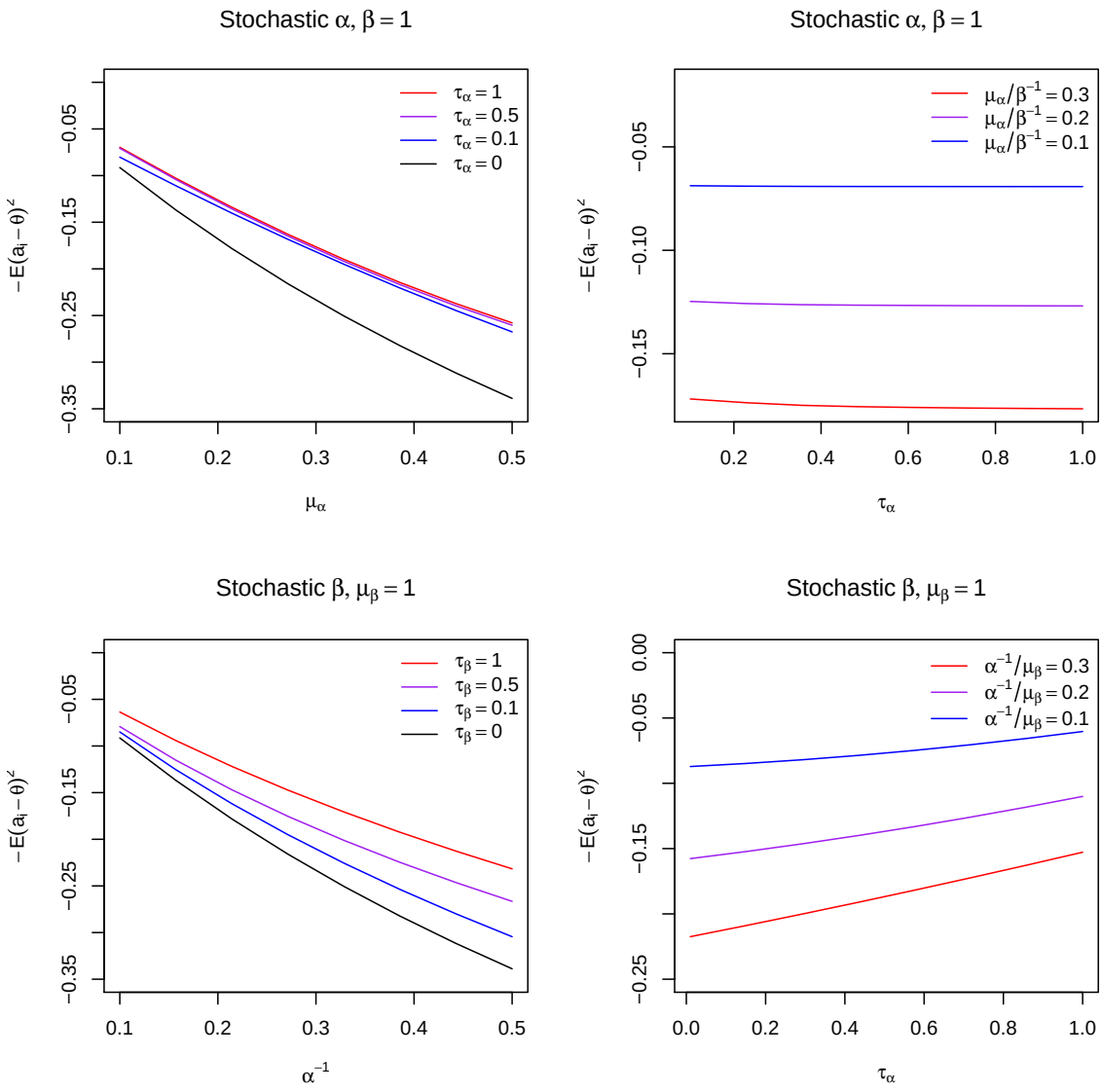


Figure 4.6: The expected welfare for $r = 0.25$ under different values of μ_α, τ_α, μ_β, and τ_β.

combinations. The expected welfare decreases with higher μ_α or μ_β. Increasing the variance-uncertainty of the public signal may increase or decrease expected welfare, and the welfare gradient is steepest for small levels of variance-uncertainty. Increasing the variance-uncertainty of the private signal increases expected welfare, and the magnitude of this effect is larger. When r is larger, expected welfare is lower.

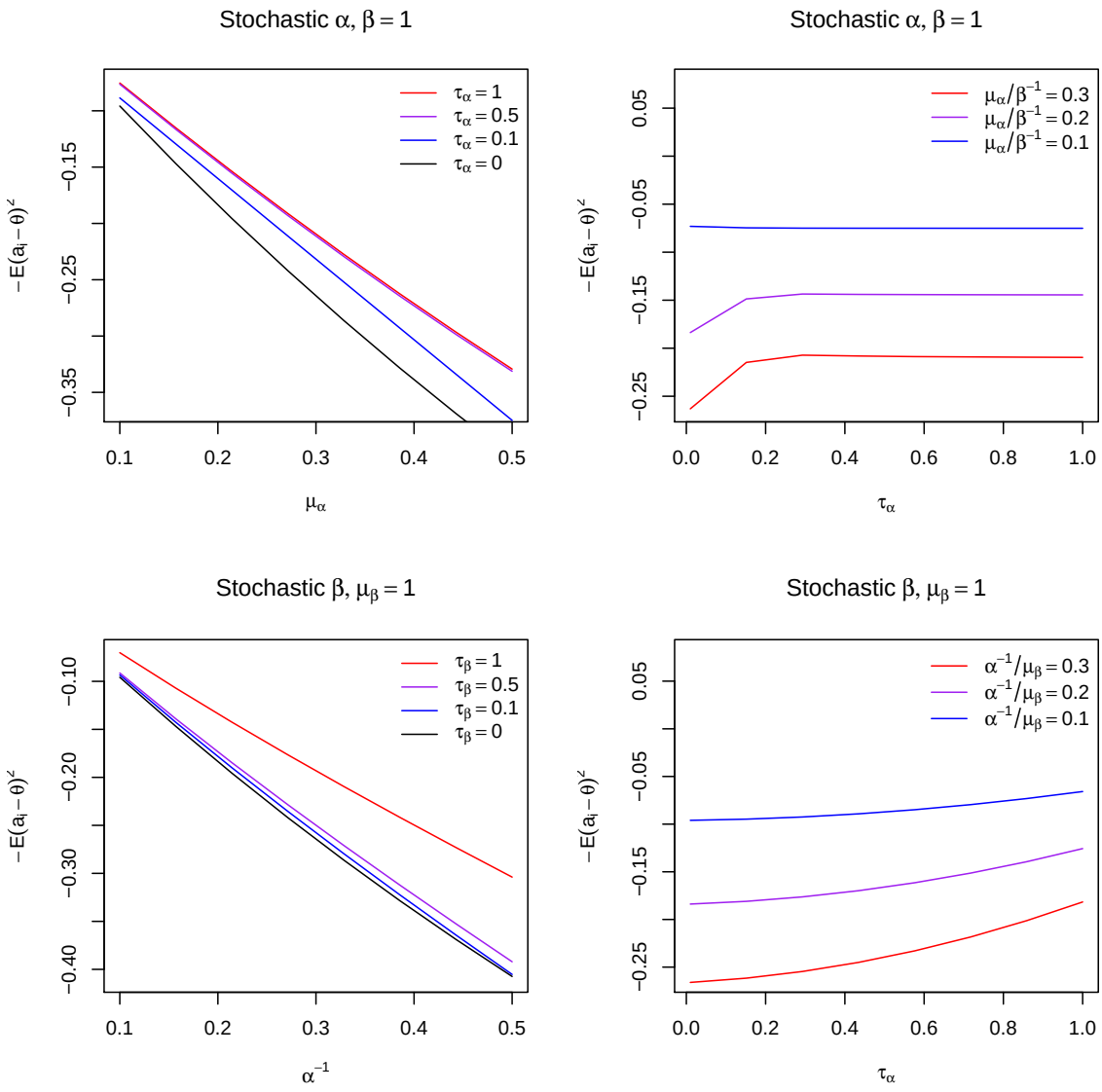


Figure 4.7: The expected welfare for $r = 0.75$ under different values of μ_α, τ_α, μ_β, and τ_β.

4.5 Conclusion

The idea of seperately modeling the effects of mean- and variance-uncertainty has applications in the literature on uncertainty shocks. Although this literature has almost always assumed that agents know the current volatility, in some cases it may be worth investigating the role of variance-uncertainty. It seems plausible that mean- and variance-uncertainty could comove. In contexts such as this model, where the two kinds of uncertainty have opposite effect signs, the failure to model

variance-uncertainty could attenuate an estimate of a mean-uncertainty effect into statistical insignificance.

One can also consider variance-uncertainty in the context of policy design. Following the publication of MS, there was some debate about its relevance for central bank transparency. MS showed that, for certain parameterizations under which agent's private signal is very precise compared to the public signal, welfare is decreasing in the precision of the public signal. This allowed for the possibility that the model supported limiting central bank transparency (Blinder et al., 2008). However, Svensson (2006) published a response to the paper that noted that this anti-transparency result only held in cases where the beauty-contest weight is high, and also when the private signal is more precise than the public signal. He argued that, in realistic central bank settings, the public signal will be more precise than the private signal, and so the model actually supports increasing central bank transparency. In an author response, Morris et al. (2006) agreed with this interpretation. This paper illustrates that the motivating intuition for this debate, that limited policy information can be beneficial, may be more applicable to variance-uncertainty.

The results from this model do not argue that a policy maker should decrease transparency. Rather, the idea of this model is that variable signal quality can be beneficial. If a policy maker is deciding between two forms of communication that have the same expected amount of noise, but one signal has constant and the other has variable quality, then this model advocates for using the signal of variable quality. This is consistent with a line of discussion in Reis (2013): he argues that there is no convincing empirical evidence for obscuring central bank information, but agent incentives and information limitations may call for optimization of the form of policy communications.

5
Conclusion

This thesis has investigated a set a models of information rigidity, with the aim of characterizing ways in which information constraints deviate from the patterns of full-information models. In this conclusion I summarize the results and discuss avenues for further work.

Chapter 2 estimated a model of forecaster behavior. The model included three principle components: an accuracy and a consistency objective, different timings for the realization of the loss associated with inaccuracy and inconsistency, and heterogeneous time discount rates. In their review of survey forecast data, Pesaran and Weale (2006) note: "To the extent that the surveys report forecasts produced as a result of some forecasting process, it is questionable how far such forecasts actually represent anyone's expectations, at least in a formal sense. Sometimes they are constructed to show what will happen if a policy which is not expected to be followed is actually followed." In this chapter, I call this excessive persistence in forecasts a consistency objective. I bracket the question of the origins of the consistency objective, but the causes of this objective is an area of further investigation. For instance, it is possible that the lack of responsiveness to new information is due to something like the variance uncertainty modelled in Chapter 4. Agents may not update their information sets because they are uncertain about the precision of their signals. Another area for further work is how heterogenous discount rates

affect expectation aggregation. Empirical work has produced very different discount rate estimates, and there is no consensus across studies as to agent's discount rates (Malhotra et al. 2002). It is possible that discount rate heterogeneity could be applied more broadly in accounting for patterns in expectation formation.

The aim of Chapter 3 was to estimate a Rational Inattention model and model inertial price-level dynamics. Although I find that the model is not capable of reproducing the patterns displayed by the data, the model is very simplified, in that it only has one aggregate shock. An important direction for further estimation work on RI models would be to create a more flexible model of aggregate dynamics. Alternatively, one could focus on a micro example where the presence of a single common shock series is more plausible.

Given these misspecification issues, the model suggested that the capacity for information processing or the cost of information has declined. This is consistent with the hypothesis that one of the contributions to the Great Moderation was the availability of higher-quality information among firms (Davis and Kahn 2008). However, there are different channels through which this information may have reduced aggregate volatility. One is through better monetary policy communication (Kahn et al. 2002). Another is better information about future demand shocks, which translates to smaller production errors, and lower output and thus price volatility (Kahn 2008). Determining which are the most important kinds of improved information flow is an area for further investigation. Understanding which information frictions have been most subject to improvement is relevant for policymakers, and also relevant for medium- to long-run forecasting.

Chapter 4 distinguishes between two kinds of uncertainty: mean-uncertainty and variance-uncertainty. It shows how they have contrasting implications in the presence of strategic complementarity. This idea could be applied in various areas: variance-uncertainty about productivity, about fiscal policy, and about monetary policy. Fiscal policy seems in particular like a significant source of variance-uncertainty for firms, which could potentially have significant output effects.

Another area of further work is the application of variance-uncertainty to models of strategic interaction. In Angeletos and Pavan (2003), for example, firms coordinate less than the social optimum. In the models in Woodford (2003), Hellwig (2005), and Roca (2010), agents have disutility from cross-sectional price dispersion. In these models, the role of variance-uncertainty would be different.

One question is how variance-uncertainty with respect to the current volatility is useful as a distinct feature from uncertainty with respect to the level of future volatility. Variance-uncertainty about the future, as with any stochastic volatility model, is a feature of Bloom (2009). A related area of investigation in finance is the work on the variance premium, defined as the difference between the option-implied (VIX) and statistical expectation of the one-month return variance. Work on this area has found that the variance premium is systematically positive (Carr and Wu 2008) and can predict stock returns (Bekaert and Hoerova 2014, Drechsler 2013). Lastly, the degree of variance-uncertainty could be measured empirically, using survey work on elicited density functions (Garthwaite et al., 2005). It is possible that empirical estimates of individual's variance-uncertainty would inform its role in forward-looking decisions.

The models in this thesis place a specific kind of information constraint on agents. There are restrictions on the kinds of signals that the agent can receive, but given this information, agents have no limits on their computational capacity to determine optimal behavior. In dividing agents into two parts — (1) a constrained information recipient and (2) an unconstrained optimizer — the models introduce a particular kind of irony. The fewer assumptions we place on agents ability to collect information, the greater the assumptions we make about their ability to perform complicated optimizing calculations. One could argue that, as public information becomes cheaper, the opposite scenario is more accurate: information is unconstrained but the time and resources required for solving optimization problems is scarce. Future work may find a way to reframe information and computational constraints in order to capture this reality.

Appendices

A

Appendix: Disagreement and consistency among professional forecasters

A.1 Stylized facts for the noisy information model

This section gives stylized facts for the model without the consistency objective.

In the noisy information model without a consistency objective, the forecasts have the form

$$y_{i,t+h|t} = \phi^{h-1} y_{i,t+1|t} \tag{A.1}$$

Consequently, the disagreement at any horizon is the scaled disagreement of the one-quarter-ahead forecast:

$$\mathrm{Var}(y_{i,t+h|t}|t) = \phi^{2(h-1)} \mathrm{Var}(y_{i,t+1|t}|t) \tag{A.2}$$

Consider the regression of one-quarter-ahead disagreement on the level of and change in inflation:

$$\mathrm{Var}(y_{i,t+1|t}|t) = \beta_0 + \beta_1 |y_t| + \beta_2 |\Delta y_t| + \text{error} \tag{A.3}$$

From (A.2), these coefficients could also be estimated using disagreement at some arbitrary horizon h:

$$\text{Var}(y_{i,t+h|t}|t) = \phi^{2(h-1)}\beta_0 + \phi^{2(h-1)}\beta_1|y_t| + \phi^{2(h-1)}\beta_2|\Delta y_t| + \text{error} \tag{A.4}$$

Therefore, as $\phi < 1$, the coefficients of a regression of disagreement on the level of and change in inflation would be decreasing in h, the horizon length. Note that this is the opposite of what is seen in Table 2.1.

Now consider the regression of one-quarter-ahead disagreement about an event on the disagreement about the same event from the previous period:

$$\text{Var}(y_{i,t+2|t+1}|t) = \gamma_0 + \gamma_1 Var(y_{i,t+2|t}|t) + \text{error} \tag{A.5}$$

From (A.2), one can rewrite this as the relationship between disagreement at any horizon and the disagreement about that event from the previous period:

$$\text{Var}(y_{i,t+h+1|t+1}|t) = \gamma_0\phi^{2(h-2)} + \gamma_1\phi^{-2}\text{Var}(y_{i,t+h|t}|t) + \text{error} \tag{A.6}$$

The coefficient on $\text{Var}(y_{i,t+h|t}|t)$ does not depend on h. The model predicts an inheritance of disagreement from one period to the next that is independent of the horizon length. This contradicts the stylized facts from Table 2.2.

A.2 Stylized facts for the consistency objective

Suppose that forecasters estimate the vector $s_{i,t|t}$ and report the vector $z_{i,t|t}$:

$$s_{i,t|t} = \begin{bmatrix} x^*_{i,t+H|t} \\ y^*_{i,t+H|t} \\ y^*_{i,t+H-1|t} \\ \vdots \\ y^*_{i,t-1|t} \end{bmatrix}, \quad z_{i,t|t} = \begin{bmatrix} \hat{x}_{i,t+H|t} \\ \hat{y}_{i,t+H|t} \\ \hat{y}_{i,t+H-1|t} \\ \vdots \\ \hat{y}_{i,t-1|t} \end{bmatrix}$$

where

$$\hat{y}_{i,t+h|t} = B_{i,h}y^*_{i,t+h|t} + (1 - B_{i,h})\hat{y}_{i,t+h|t-1}$$

From the Kalman filter, note that $s_{i,t|t} = (1 - K_t H)Fs_{i,t-1|t-1} + K_t H s_t + K_t w_{i,t}$. In steady state, K_t converges to K. Let $k_{h,l}$ refer to a row-column entry of K. The update of a single entry in the state can be written:

$$y^*_{i,t+h|t} = k_{h,2}w_{i,t} + k_{h,2}y_t + k_{h,1}y_{t-1} + y^*_{i,t+h|t-1} - k_{h,2}y^*_{i,t|t-1} - k_{h,1}y^*_{i,t-1|t-1} \quad \text{(A.7)}$$

Consequently,

$$\hat{y}_{i,t+h|t} = B_{i,h}(k_{h,2}w_{i,t} + k_{h,2}y_t + k_{h,1}y_{t-1} + y^*_{i,t+h|t-1} - k_{h,2}y^*_{i,t|t-1} - k_{h,1}y^*_{i,t-1|t-1}) + (1 - B_{i,h})\hat{y}_{i,t+h|t-1} \quad \text{(A.8)}$$

Equivalently,

$$\hat{y}_{i,t+h|t} = B_{i,h}(k_{h,2}w_{i,t} + (k_{h,2} + k_{h,1})y_t - k_{h,1}\Delta y_t + (\phi^{h-1} - k_{h,2})y^*_{i,t|t-1} - k_{h,1}y^*_{i,t-1|t-1}) + (1 - B_{i,h})\hat{y}_{i,t+h|t-1}$$

Consequently,

$$\begin{aligned}
\mathrm{Var}(\hat{y}_{i,t+h|t}|t, y_t, \Delta y_t) = {} & \mathrm{Var}(B_{i,h}k_{h,2}w_{i,t}) \\
& + \mathrm{Var}(B_{i,h}(k_{h,2} + k_{h,1}))y_t^2 \\
& + \mathrm{Var}(B_{i,h}k_{h,1})(\Delta y_t)^2 \\
& + \mathrm{Var}(B_{i,h}(\phi^{h-1} - k_{h,2})y^*_{i,t|t-1}) \\
& + \mathrm{Var}(B_{i,h}k_{h,1}y^*_{i,t-1|t-1}) \\
& + \mathrm{Var}((1 - B_{i,h})\hat{y}_{i,t+h|t-1}) \\
& + Cov(B_{i,h}(\phi^{h-1} - k_{h,2})y^*_{i,t|t-1}, B_{i,h}k_{h,1}y^*_{i,t-1|t-1}) \\
& + Cov(B_{i,h}(\phi^{h-1} - k_{h,2})y^*_{i,t|t-1}, (1 - B_{i,h})\hat{y}_{i,t+h|t-1}) \\
& + Cov(B_{i,h}k_{h,1}y^*_{i,t-1|t-1}, (1 - B_{i,h})\hat{y}_{i,t+h|t-1})
\end{aligned}$$

This is increasing in y_t^2, Δy_t^2, and the slope depends on $\mathrm{Var}(B_{i,h})$. That is, the greater the heterogeneity in the weights B_i, the stronger the relationship between disagreement, y_t^2, and Δy_t^2.

A.3 GMM moments

A.3.1 Inflation process moments

Three moments were used for the inflation process itself:

$$Cov(y_t, y_{t-1}) = \phi$$

$$\text{Var}(y_t) = \frac{\sigma_\epsilon^2}{1-\phi^2} + \sigma_u^2$$

$$E((y_t - \phi y_{t-1})y_{t-2}) = 0$$

A.3.2 Consistency parameter moments

Moments were constructed from the condition that the state estimates are uncorrelated with the (1) level and (2) change in the nowcast. These two sets of conditions are constructed in the same way.

Begin with the condition that the forecast error is uncorrelated with the change in inflation:

$$E\left((y^*_{i,t+h|t} - y_{t+h})(\Delta y^*_{i,t|t})\right) = 0$$

$$E\left[\left(\frac{1}{B_{i,h}}\hat{y}_{i,t+h|t} + \left(1 - \frac{1}{B_{i,h}}\right)\hat{y}_{i,t+h|t-1} - y_{t+h}\right)(\Delta y^*_{i,t|t})\right] = 0$$

$$E\left[\left(\frac{1}{B_{i,h}}\left(\hat{y}_{i,t+h|t} - \hat{y}_{i,t+h|t-1}\right) + \hat{y}_{i,t+h|t-1} - y_{t+h}\right)(\Delta y^*_{i,t|t})\right] = 0$$

Combining horizons and time periods:

$$E\left[\left(M_i\left(\hat{Y}'_{i,0} - \hat{Y}'_{i,-1}\right) + \hat{Y}'_{i,-1} - Y'_T\right)(\Delta Y^*_T)\right] = 0$$

Where

$$M_i = \begin{bmatrix} B_{i,4}^{-1} & 0 & 0 & 0 \\ 0 & B_{i,3}^{-1} & 0 & 0 \\ 0 & 0 & B_{i,2}^{-1} & 0 \\ 0 & 0 & 0 & B_{i,1}^{-1} \end{bmatrix}, \hat{Y}_{i,0} = \begin{bmatrix} \hat{y}_{i,5|1} & \hat{y}_{i,4|1} & \hat{y}_{i,3|1} & \hat{y}_{i,2|1} \\ \hat{y}_{i,6|2} & \hat{y}_{i,5|2} & \hat{y}_{i,4|2} & \hat{y}_{i,3|2} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{y}_{i,T+4|T} & \hat{y}_{i,T+3|T} & \hat{y}_{i,T+2|T} & \hat{y}_{i,T+1|T} \end{bmatrix},$$

$$\hat{Y}_{i,-1} = \begin{bmatrix} \hat{y}_{i,5|0} & \hat{y}_{i,4|0} & \hat{y}_{i,3|0} & \hat{y}_{i,2|0} \\ \hat{y}_{i,6|1} & \hat{y}_{i,5|1} & \hat{y}_{i,4|1} & \hat{y}_{i,3|1} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{y}_{i,T+4|T-1} & \hat{y}_{i,T+3|T-1} & \hat{y}_{i,T+2|T-1} & \hat{y}_{i,T+1|T-1} \end{bmatrix},$$

$$Y_T = \begin{bmatrix} \hat{y}_{i,5} & \hat{y}_{i,4} & \hat{y}_{i,3} & \hat{y}_{i,2} \\ \hat{y}_{i,6} & \hat{y}_{i,5} & \hat{y}_{i,4} & \hat{y}_{i,3} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{y}_{i,T+4} & \hat{y}_{i,T+3} & \hat{y}_{i,T+2} & \hat{y}_{i,T+1} \end{bmatrix}, \Delta Y_T = \begin{bmatrix} (\Delta y_{i,1|1}^*) \\ (\Delta y_{i,2|2}^*) \\ \vdots \\ (\Delta y_{i,T|T}^*) \end{bmatrix}.$$

We then have a moment condition for each agent.

These moments can also be produced for the other condition, that the forecast error is uncorrelated with the level, in an analogous way:

$$E\left((y_{i,t+h|t}^* - y_{t+h})(y_{i,t|t}^*)\right) = 0$$

A.3.3 Cross-sectional variance moments

The final moment conditions are from the cross-sectional variance of the state estimate: $\text{Var}(y_{i,t|t-h}^*|t)$.

The sample moments come from transforming the reported forecasts using B and β_i to get the implied state estimates, $y_{i,t|t-h}^*$. Then one can take the sample estimate:

$$\text{E}(Var(y_{i,t|t-h}^*|t)) = P_{t|t},\ h = 1, 2, 3, 4$$

The population moment can be approximated with the Kalman filter. The steady state value of $P_{t|t}$ can be obtained by iterating the variance updating formulas from the filter 100 times:

1. Begin with symmetric positive definite $P_{i,t-1|t-1}$
2. $P_{i,t|t-1} = FP_{i,t-1|t-1}F' + Q$
3. $\Sigma_{i,t|t-1} = HP_{i,t|t-1}H' + R.$
4. $K_{i,t} = P_{i,t|t-1}H'\Sigma^{-1}_{i,t|t-1}.$
5. $P_{i,t|t} = (I - K_{i,t})P_{i,t|t-1}.$

B

Appendix: Estimation of a Rational Inattention model with multiple volatility regimes

B.1 Proof of Proposition 3.1

Suppose an RI agent must report an estimate $\hat{x}$ of a variable x. x is a known linear combination of jointly Gaussian variables, x_1 and x_2. Suppose that the loss function ℓ is quadratic with some diagonal positive weighting matrix W: $\ell \equiv -E((x-\hat{x})'W(x-\hat{x}))$. Further suppose that x_1 and x_2 are independent in the agent's prior distribution:

$$(x_1, x_2)' \sim N((\mu_1, \mu_2)', diag(\sigma_1^2, \sigma_2^2)). \tag{B.1}$$

Consider an arbitrary 2×1 vector signal $\hat{x}_{1;2}$ in which both entries have nonzero covariance with both x_1 and x_2. Then there will exist a 2×1 vector signal, $(\hat{x}_1, \hat{x}_2)$, in which $Cov(\hat{x}_1, \hat{x}_2) = 0$, such that the agent can achieve the same loss with $(\hat{x}_1, \hat{x}_2)$ as with $\hat{x}_{1;2}$.

Proof:

Consider the 2×1 vector signal $\hat{x}_{1;2}$. The signal $\hat{x}_{1;2}$ will be a bijective linear transformation of $\hat{x}$:

$$\hat{x}_{1;2} = B\hat{x} \tag{B.2}$$

The information constraint is:

$$I(\tilde{x}_{1;2}, \hat{x}_{1;2}) \leq \kappa \tag{B.3}$$
$$\frac{\det \text{Var}_{prior}(\tilde{x}_{1;2})}{\det \text{Var}_{post}(\tilde{x}_{1;2})} \leq \exp(2\kappa) \tag{B.4}$$
$$\frac{\det(B\text{Var}_{prior}(\tilde{x})B')}{\det(B\text{Var}_{post}(\tilde{x})B')} \leq \exp(2\kappa) \tag{B.5}$$
$$\frac{\det B \det \text{Var}_{prior}(\tilde{x}) \det B'}{\det B \det \text{Var}_{post}(\tilde{x}) \det B'} \leq \exp(2\kappa) \tag{B.6}$$
$$\frac{\det \text{Var}_{prior}(\tilde{x})}{\det \text{Var}_{post}(\tilde{x})} \leq \exp(2\kappa) \tag{B.7}$$

The same calculation holds for the signal $(\hat{x}_1, \hat{x}_2)$. Therefore, the two signals $\hat{x}_{1;2}$ and $(\hat{x}_1, \hat{x}_2)$ are equivalent in the sense that the agent will be able to achieve the same loss with some signal $(\hat{x}_1, \hat{x}_2)$ as with $\hat{x}_{1;2}$. This proof simply restates the fact that mutual information is invariant to linear transformations. ∎

B.2 Proof of Proposition 3.2

Consider an RI agent as described in (3.19) - (3.23). The agent will choose to receive independent signals of each innovation $\epsilon_{z,t-l}$, $l \geq 0$. *That is,* $\tilde{z}_t = (\epsilon_{z,t}, \epsilon_{z,t-1}, ..., \epsilon_{z,-\infty})'$.

Proof: From the result in (3.25), we know that the agent will choose a Gaussian signal. We know that it will be possible to write $\tilde{z}$ as a linear combination of the history of $\epsilon_{z,t}$:

$$\tilde{z} = H_z(\epsilon_{z,t}, \epsilon_{z,t-1}, ..., \epsilon_{z,-\infty})' \tag{B.8}$$

where H_z is a real-valued matrix. From repeated application of Proposition 3.1, we know that each entry in $\tilde{z}$ will depend on at most one innovation of z_t. It is not possible that an entry in $\tilde{z}$ will depend on 0 innovations of z_t, so H_z has exactly one nonzero entry per row. It is not possible for the agent to receive multiple

signals of the same innovation as this would violate the requirement in the RI model specification that the prior variance be positive definite, as in (3.7).

Note that the agent's posterior will be invariant to certain rescalings of the signal choice. In (B.8), if D is any diagonal matrix, we can premultiply H_z by D and it will produce the same posterior distribution for $z_{t|t}$. Therefore we can restrict attention to the case where H_z is composed of 0s and 1s without loss of generality. Then $\tilde{z} = (\epsilon_{z,t}, \epsilon_{z,t-1}, ..., \epsilon_{z,-\infty})'$. ∎

B.3 Proof of Proposition 3.3

Consider an RI agent tracking z_t as described in (3.19) - (3.23). Suppose there are initial conditions such that in period t it is optimal for the agent to pay non-zero attention to (at least one) lagged innovation $\epsilon_{z,t-k}$:

$$\frac{Var(\epsilon_{z,t-k}|\mathcal{I}_{t-1})}{Var(\epsilon_{z,t-k}|\mathcal{I}_t)} > 1,\ k > 0 \tag{B.9}$$

Within a finite number of periods $j^* < \infty$, the agent will pay zero attention to all lagged innovations:

$$\frac{Var(\epsilon_{z,t+j-k}|\mathcal{I}_{t+j-1})}{Var(\epsilon_{z,t+j-k}|\mathcal{I}_{t+j})} = 1\ \forall j > j^*, k > 0 \tag{B.10}$$

Proof: This is a special case of Proposition 3.6, proved in Appendix B.6, where the attention allocation to Δ_t is 0. ∎

B.4 Proof of Proposition 3.4

Consider an RI agent as described in (3.19) - (3.23). Consider the limiting loss of a signal strategy, $\ell_t(\hat{z}_t)|\mathcal{I}_t$ as $t \to \infty$. The agent will minimize this loss by always choosing a scalar signal of the current innovation, ie $\tilde{z}_t = \epsilon_{z,t}$:

$$\arg\left(\min_{\tilde{z}_t}\left(\lim_{t\to\infty}\sum_{s=t}^{\infty}\beta^{s-t}\sigma^2_{s|s}\right)\right) = \epsilon_{z,t} \tag{B.11}$$

Proof: Let $\text{Var}(\epsilon_{z,t}|\mathcal{I}_s)$ be denoted $\tau^2_{z,t|s}$. Note that the marginal variance of the innovations is $\tau^2_{z,t|t-1} = Var(\epsilon_{z,t}) = \phi^2_z$.

Given Proposition 3.2, the optimal signal will be of the form $\tilde{z} = (\epsilon_{z,t}, \epsilon_{z,t-1}, ..., \epsilon_{z,-\infty})'$. Below I compare two scenarios:

1. Strictly positive attention at time t is allocated to $\epsilon_{z,t}$ and $\epsilon_{z,t-1}$
2. Strictly positive attention is only allocated to $\epsilon_{z,t}$

(1) Suppose strictly positive attention is allocated to $\epsilon_{z,t}$ and $\epsilon_{z,t-1}$. The information constraint is:

$$\frac{\phi_z^2}{\tau_{z,t|t}^2}\frac{\tau_{z,t-1|t-1}^2}{\tau_{z,t-1|t}^2} = \exp(2\kappa_z) \tag{B.12}$$

The assumption of strictly positive attention implies that $\frac{\tau_{z,t-1|t-1}^2}{\tau_{z,t-1|t}^2} = \rho > 1$. Consequently,

$$\tau_{z,t|t}^2 = \rho\frac{\phi_z^2}{\exp(2\kappa_z)} \tag{B.13}$$

Agents remember information gathered in the previous period, hence in steady state $\tau_{z,t|t}^2 = \tau_{z,t-1|t-1}^2$. Consequently,

$$\tau_{z,t-1|t}^2 = \frac{\phi_z^2}{\exp(2\kappa_z)} \tag{B.14}$$

(2) Now suppose no attention is paid to $\epsilon_{z,t-1}$. Then $\frac{\sigma_{z,t-1|t-1}^2}{\sigma^2 z,t-1|t} = 1$,

$$\tau_{z,t|t}^2 = \frac{\phi_z^2}{\exp(2\kappa_z)} \tag{B.15}$$

and

$$\tau_{z,t-1|t}^2 = \frac{\phi_z^2}{\exp(2\kappa_z)}. \tag{B.16}$$

Comparison

The loss is of the form $\alpha_{z,0}^2\tau_{z,t|t}^2 + \alpha_{z,1}^2\tau_{z,t-1|t}^2 + \sum_{l=2}^{\infty}\alpha_{z,l}^2\tau_{z,t-l|t}^2$. We can compute the loss for (1) and (2):

$$\ell_{(1)} = \frac{\sigma^2}{\exp(2\kappa_z)} \left(\alpha_0 \rho + \alpha_1 + \sum_{l=2}^{\infty} \alpha_{z,l}^2 \tau_{z,t-l|t}^2 \right) \tag{B.17}$$

$$\ell_{(2)} = \frac{\sigma^2}{\exp(2\kappa_z)} \left(\alpha_0 + \alpha_1 + \sum_{l=2}^{\infty} \alpha_{z,l}^2 \tau_{z,t-l|t}^2 \right) \tag{B.18}$$

Since $\rho > 1$, $\ell_{(1)} > \ell_{(2)}$. We can conclude that it is optimal to allocate all attention to $\epsilon_{z,t}$.

This logic yields the result that any attention devoted to a lagged innovation $\epsilon_{z,t-l}$ will optimally be be reallocated to $\epsilon_{z,t-l+1}$. Repeated allocation of this argument yields the result that all attention is optimally allocated to the current innovation:

$$\arg \left(\min_{\tilde{z}_t} \left(\lim_{t \to \infty} \sum_{s=t}^{\infty} \beta^{s-t} \sigma_{s|s}^2 \right) \right) = \epsilon_{z,t} \tag{B.19}$$

∎

B.5 Proof of Proposition 3.5

Consider an RI agent as described in (3.19) - (3.23) who uses Kalman filtering. Suppose following Proposition 3.3 she pays attention only to ϵ_{zt} in period t, $\forall t$. She tracks the variable

$$z_t = \sum_{l=0}^{\infty} \alpha_{z,l} \epsilon_{z,t-l} \tag{B.20}$$

where $\epsilon_{z,t} \sim N(0, \phi_z^2)$. Then the agent's response function $z_{t|t}$ will have the following form:

$$z_{t|t} = \sum_{l=0}^{\infty} \alpha_{z,l} \epsilon_{z,t-1|t-l} \tag{B.21}$$

where $\epsilon_{z,t|t} \sim N(0, \phi_z^2)$ and

$$Cor(z_t, z_{t|t}) = \frac{1}{1 + (\exp(2\kappa_z) - 1)^{-1}} \tag{B.22}$$

Proof: At time t, the agent performs the following update:

1. The prior distribution is $\epsilon_{z,t} \sim N(0, \phi_z^2)$

2. The agent chooses a noise variance ψ_z^2 to satisfy the information constraint: $\text{Var}(\epsilon_t)/\text{Var}(\epsilon_t|\hat{\epsilon}_{z,t}) = 1 + \frac{\phi_z^2}{\psi_z^2} = \exp(2\kappa_z)$

3. The signal is distributed $\hat{\epsilon}_{z,t} \sim N(\epsilon_{z,t}, \psi_z^2)$

4. The Kalman gain is $\phi_z^2/(\phi_z^2 + \psi_z^2)$

5. The posterior estimate is $\epsilon_{z,t|t} = \phi_z^2\hat{\epsilon}_{z,t}/(\phi_z^2 + \psi_z^2)$

6. The marginal distribution of the posterior estimate is $\epsilon_{z,t|t} \sim N(0, \phi_z^2)$

7. The agent forms her estimate $z_{t|t} = \sum_{l=0}^{\infty} \alpha_{z,l}\epsilon_{t-l|t-l}$

Then $\text{Cov}(z_{t|t}, z_{t-k|t-k}) = \text{Cov}(z_t, z_{t-k})\forall t, k$. Hence

$$z_{t|t} = \sum_{l=0}^{\infty} \alpha_{z,l}\epsilon_{z,t-1|t-l} \tag{B.23}$$

where $\epsilon_{z,t|t} \sim N(0, \phi_z^2)$.

Moreover, $\epsilon_{z,t|t} = \phi_z^2\hat{\epsilon}_{z,t}/(\phi_z^2+\psi_z^2)$, and combined with the information constraint $1 + \frac{\phi_z^2}{\psi_z^2} = \exp(2\kappa_z)$, yields:

$$\text{Cor}(z_t, z_{t|t}) = \frac{1}{1 + (\exp(2\kappa_z) - 1)^{-1}} \tag{B.24}$$

Which are the desired results. ∎

B.6 Proof of Proposition 3.6

Consider an RI agent as described in (3.34) - (3.38). Suppose there are initial conditions such that in period t it is optimal for the agent to pay non-zero attention to (at least one) lagged innovation $\epsilon_{i,t-k}$:

$$\frac{Var(\epsilon_{i,t-k}|\mathcal{I}_{t-1})}{Var(\epsilon_{i,t-k}|\mathcal{I}_t)} > 1,\ k > 0,\ i \in \{\Delta, z\} \tag{B.25}$$

Within a finite number of periods $j^* < \infty$, the agent will pay zero attention to all lagged innovations:

$$\frac{Var(\epsilon_{i,t+j-k}|\mathcal{I}_{t+j-1})}{Var(\epsilon_{z,t+j-k}|\mathcal{I}_{t+j})} = 1\ \forall j > j^*, k > 0, i \in \{\Delta, z\} \tag{B.26}$$

Proof:

I show that if, in period t, the agent pays attention just to the innovations $\{\epsilon_t, \epsilon_{t-1}\}$, after a finite number of periods the agent will only pay attention to ϵ_t. This argument can be repeatedly applied for higher order lags to achieve the desired result.

Propositions (3.1 - 3.2) hold, so $\tilde{z}_t = (\epsilon_{z,t}, \epsilon_{z,t-1}, ...)$.

Introduce the following notation:

$$\text{Var}(\epsilon_t|\mathcal{I}_s) = \tau^2_{t|s} \tag{B.27}$$

We can write the information constraint in terms of the attention paid to each innovation:

$$\prod_{l=0}^{L} \frac{\tau^2_{\Delta,t-l|t-1}}{\tau^2_{\Delta,t-l|t}} \prod_{l=0}^{L} \frac{\tau^2_{z,t-l|t-1}}{\tau^2_{z,t-l|t}} = \kappa^* \tag{B.28}$$

where $\kappa^* = \exp(2\kappa)$ And we can write the loss function similarly:

$$\ell_t = \sum_{s=t}^{\infty} \beta^{s-t} \left(\sum_{l=0}^{L} \alpha^2_{\Delta,l} \tau^2_{\Delta,s-l|s} + \sum_{l=0}^{L} \alpha^2_{z,l} \tau^2_{z,s-l|s} \right) \tag{B.29}$$

Plug the constraint into the loss:

$$\ell_t = \alpha^2_{\Delta,0} \frac{\tau^2_{\Delta,t|t-1}}{\kappa^*} \prod_{l=0}^{L} \frac{\tau^2_{z,t-l|t-1}}{\tau^2_{z,t-l|t}} + \tag{B.30}$$

$$\left(\sum_{l=1}^{L} \alpha^2_{\Delta,l} \tau^2_{\Delta,t-l|t} + \sum_{l=0}^{L} \alpha^2_{z,l} \tau^2_{z,t-l|t} \right) + \tag{B.31}$$

$$\sum_{s=t+1}^{\infty} \beta^{s-t} \left(\sum_{l=0}^{L} \alpha^2_{\Delta,l} \tau^2_{\Delta,s-l|s} + \sum_{l=0}^{L} \alpha^2_{z,l} \tau^2_{z,s-l|s} \right) \tag{B.32}$$

Now suppose it is optimal at time t to pay nonzero attention to $\epsilon_{\Delta,t}, \epsilon_{z,t}, \epsilon_{z,t-1}$.

The derivatives with respect to $\tau^2_{z,t|t}$ and $\tau^2_{z,t-1|t}$ are:

$$\frac{\partial \ell_t}{\partial \tau^2_{z,t|t}} = -\alpha^2_{\Delta,0} \frac{\tau^2_{\Delta,t|t-1}}{\kappa^*} \frac{\phi^2_{z,t}}{\tau^4_{z,t|t}} \frac{\tau^2_{z,t-1|t-1}}{\tau^2_{z,t-1|t}} + \alpha^2_{z0} \tag{B.33}$$

$$+ \frac{\alpha^2_{\Delta,0}}{\tau^2_{z,t|t}} \frac{\tau^2_{\Delta,t+1|t}}{\kappa^*} \prod_{l=0}^{L} \frac{\tau^2_{z,t-l|t}}{\tau^2_{z,t-l|t+1}} \tag{B.34}$$

$$\frac{\partial \ell_t}{\partial \tau^2_{z,t-1|t}} = -\alpha^2_{\Delta,0} \frac{\tau^2_{\Delta,t|t-1}}{\kappa^*} \frac{\phi^2_{z,t}}{\tau^2_{z,t|t}} \frac{\tau^2_{z,t-1|t-1}}{\tau^4_{z,t-1|t}} + \alpha^2_{z1} \tag{B.35}$$

$$+ \frac{\alpha^2_{\Delta,0}}{\tau^2_{z,t-1|t}} \frac{\tau^2_{\Delta,t+1|t}}{\kappa^*} \prod_{l=0}^{L} \frac{\tau^2_{z,t-l|t}}{\tau^2_{z,t-l|t+1}} \tag{B.36}$$

The agent will always pay strictly positive attention to $\epsilon_{z,t}$ at time t, because $|\alpha_{z,l}| > |\alpha_{z,l+1}| \forall l \geq 0$.

We have posited that the following is true at time t:

$$\frac{\partial \ell_t}{\partial \tau^2_{z,t|t}} = \frac{\partial \ell_t}{\partial \tau^2_{z,t-1|t}} \tag{B.37}$$

Consequently,

$$\frac{(\alpha^2_{z0} - \alpha^2_{z1})\kappa^*}{\alpha^2_{\Delta,0}\phi^2_{\Delta}} = \left(\frac{1}{\tau^2_{z,t-1|t}} - \frac{1}{\tau^2_{z,t|t}} \right) \left(\prod_{l=0}^{L} \frac{\tau^2_{z,t-l|t}}{\tau^2_{z,t-l|t+1}} - \frac{\phi^2_{z,t}}{\tau^2_{z,t|t}} \frac{\tau^2_{z,t-1|t-1}}{\tau^2_{z,t-1|t}} \right) \tag{B.38}$$

Note the assumption that $|\alpha_{z,l}| > |\alpha_{z,l+1}| \forall l \geq 0$. Then, showing that $\tau^2_{t|t} < \tau^2_{t-1|t}$ is equivalent to showing that the right hand term in (B.38) is negative, ie that

$$\prod_{l=0}^{L} \frac{\tau^2_{z,t-l|t}}{\tau^2_{z,t-l|t+1}} < \frac{\phi^2_{z,t}}{\tau^2_{z,t|t}} \frac{\tau^2_{z,t-1|t-1}}{\tau^2_{z,t-1|t}} \tag{B.39}$$

This will hold because decreasing attention will be paid to a given innovation in each subsequent period:

$$\frac{\tau^2_{z,t|t}}{\tau^2_{z,t|t+1}} < \frac{\phi^2_{z,t}}{\tau^2_{z,t|t}} \tag{B.40}$$

Therefore, $\tau^2_{t|t} < \tau^2_{t-1|t}$. Moreover, $\tau^2_{t|t} < \tau^2_{t-1|t} < \tau^2_{t-1|t-1}$, and so $\tau^2_{t|t} < \tau^2_{t-1|t-1}$.

Moreover, we know that since increasing attention will be paid to the current innovation, the rightmost term is negative and increasing as $t \to \infty$. Consequently the following term is negative and decreasing as $t \to \infty$:

$$\left(\frac{1}{\tau^2_{z,t-1|t}} - \frac{1}{\tau^2_{z,t|t}} \right) \tag{B.41}$$

Which implies that

$$\frac{\tau^2_{z,t-1|t-1}}{\tau^2_{z,t-1|t}} \to 1 \text{ as } t \to \infty \tag{B.42}$$

The agent will stop paying attention to the lagged innovation after a finite number of periods because they will encounter a corner solution. That is, there is no solution to the first order condition of the interior problem when $\tau^2_{z,t-1|t-1} = \tau^2_{z,t-1|t}$. That is to say, the following system has no solution:

$$\frac{\partial \ell_t}{\partial \tau^2_{z,t|t}} = \frac{\partial \ell_t}{\partial \tau^2_{z,t-1|t}} \text{ and } \tau^2_{z,t|t} = \tau^2_{z,t-1|t} \tag{B.43}$$

∎

B.7 Proof of Proposition 3.7

.

Consider an RI agent as described in (3.34) - (3.38). Consider the limiting loss of a signal strategy, $\ell_t|\mathcal{I}_t$ as $t \to \infty$. The agent will minimize this loss by always choosing a scalar signal of the current innovations:

$$\arg\left(\min_{\tilde{z}_t} \left(\lim_{t\to\infty} \sum_{s=t}^{\infty} \beta^{s-t} \sigma^2_{s|s} \right) \right) = (\epsilon_{\Delta,t}, \epsilon_{z,t})' \tag{B.44}$$

Proof: Fix some history of attention allocation, $\{\kappa_{\Delta,t}, \kappa_{z,t}, \kappa_{\Delta,t-1}, \kappa_{z,t-1}, ...\}$. Given this allocation, Proposition 3.4 holds for both Δ_t and z_t. Since it holds for all attention allocations, it will hold for the optimal allocation. ∎

B.8 Proof of Proposition 3.8

Suppose we observe stationary Gaussian data that consist of the following parameters from p_t *and* Δ_t*:*

$$p_t = \sum_{l=0}^{\infty} \alpha_{p,l} \epsilon_{p,t-l},\ \epsilon_{p,t} \sim N(0, \phi_p^2) \tag{B.45}$$

$$\Delta_t = \sum_{l=0}^{\infty} \alpha_{\Delta,l} \epsilon_{\Delta,t-l},\ \epsilon_{p,t} \sim N(0, \phi_\Delta^2) \tag{B.46}$$

$$\rho = Cor(p_t, \Delta_t) > 0 \tag{B.47}$$

Then there exist unique attention parameters κ_Δ *and* κ_z *that would produce the given parameters from the RI specification in (3.34) - (3.38).*

If $\rho = Cor(p_t, \Delta_t) = 0$*, then* $\kappa_z = \kappa$ *and it is only possible to set-identify* κ*. That is, it is possible to find a finite bound* $\bar{\kappa}$ *such that* $0 \leq \kappa < \bar{\kappa}$*.*

Proof: The identification is sketched in the following steps.

(1) Suppose $\text{Cor}(p_t, \Delta_t) > 0$.

The parameters of Δ_t and p_t are given.

The autocovariances of z_t are given by

$$Cov(p_t, p_{t-k}) = Cov(\Delta_t, \Delta_{t-k}) + Cov(z_t, z_{t-k}) \tag{B.48}$$

The parameters of z_t are identified by the autocovariances.

The following relationship holds and identifies ψ_Δ^2:

$$p_t = \frac{\phi_\Delta^2}{\phi_\Delta^2 + \psi_\Delta^2} \Delta_t + (\text{terms independent of } \Delta_t) \tag{B.49}$$

The information constraint identifies κ_Δ:

$$1 + \frac{\phi_\Delta^2}{\psi_\Delta^2} = \exp(2\kappa_\Delta) \tag{B.50}$$

The optimality condition identifies κ_z:

$$\sum_{s=t}^{\infty} \beta^{s-t} \alpha_{s-t}^2 \phi_\Delta^2 \exp(-2\kappa_\Delta) = \sum_{s=t}^{\infty} \beta^{s-t} \alpha_{s-t}^2 \phi_z^2 \exp(-2\kappa_z) \tag{B.51}$$

(2) Suppose $\text{Cor}(p_t, \Delta_t) = 0$. Then the agent is paying 0 attention to Δ_t. The estimate will be 0 in every period: $\Delta_{t|t} = 0 \ \forall t$

The parameters of Δ_t and p_t are given. The autocovariances of z_t are given by $Cov(p_t, p_{t-k}) = Cov(z_t, z_{t-k})$. The parameters of z_t are identified by the autocovariances. The optimality condition gives the criterion for the corner solution which bounds $\kappa = \kappa_z$:

$$\sum_{s=t}^{\infty} \beta^{s-t} \alpha_{s-t}^2 \phi_{\Delta}^2 < \sum_{s=t}^{\infty} \beta^{s-t} \alpha_{s-t}^2 \phi_z^2 \exp(-2\kappa) \tag{B.52}$$

∎

B.9 Personal Consumption Expenditure goods indices

	PCE Label	Description
1	P1NDCG3	New domestic autos
2	P1NFCG3	New foreign autos
3	P1NETG3	Net transactions in used autos
4	P1MARG3	Used auto margin
5	P1TRUG3	Trucks, new and net used
6	P1REVG3	Recreational vehicles
7	P1TATG3	Tires and tubes
8	P1PAAG3	Accessories and parts
9	P1FNRG3	Furniture, including mattresses and bedsprings
10	P1MHAG3	Major household appliances
11	P1SEAG3	Small electric appliances
12	P1CHNG3	China, glassware, tableware, and utensils
13	P1RADG3	Video and audio goods, including musical instruments, and computer goods
14	P1FLRG3	Floor coverings
15	P1CLFG3	Clocks, lamps, and furnishings
16	P1TEXG3	Blinds, rods, and other
17	P1WTRG3	Writing equipment
18	P1HDWG3	Tools, hardware, and supplies
19	P1LWNG3	Outdoor eqpt and supplies
20	P1OPTG3	Ophthalmic products and orthopedic appliances
21	P1GUNG3	Guns
22	P1SPTG3	Sporting equipment
23	P1CAMG3	Photographic equipment
24	P1BCYG3	Bicycles
25	P1MCYG3	Motorcycles
26	P1BOAG3	Pleasure boats
27	P1AIRG3	Pleasure aircraft
28	P1JRYG3	Jewelry and watches
29	P1BKSG3	Books and maps
30	P1GRAG3	Cereals
31	P1BAKG3	Bakery products
32	P1BEEG3	Beef and veal
33	P1PORG3	Pork
34	P1MEAG3	Other meats
35	P1POUG3	Poultry
36	P1FISG3	Fish and seafood
37	P1GGSG3	Eggs
38	P1MILG3	Fresh milk and cream
39	P1DAIG3	Processed dairy products
40	P1FRUG3	Fresh fruits
41	P1VEGG3	Fresh vegetables
42	P1PFVG3	Processed fruits and vegetables
43	P1JNBG3	Juices and nonalcoholic drinks
44	P1CTMG3	Coffee, tea and beverage materials
45	P1FATG3	Fats and oils

	PCE Label	Description
46	P1SWEG3	Sugar and sweets
47	P1OFDG3	Other foods
48	P1PEFG3	Pet food
49	P1MLTG3	Beer and ale, at home
50	P1WING3	Wine and brandy, at home
51	P1LIQG3	Distilled spirits, at home
52	P1ESLG3	Elementary and secondary school lunch
53	P1HSLG3	Higher education school lunch
54	P1OPMG3	Other purchased meals
55	P1APMG3	Alcohol in purchased meals
56	P1CFDG3	Food supplied civilians
57	P1MFDG3	Food supplied military
58	P1FFDG3	Food produced and consumed on farms
59	P1SHUG3	Shoes (12)
60	P1WGCG3	Clothing for females
61	P1WICG3	Clothing for infants
62	P1WSGG3	Sewing goods for females
63	P1WUGG3	Luggage for females
64	P1MBCG3	Clothing for males
65	P1MSGG3	Sewing goods for males
66	P1MUGG3	Luggage for males
67	P1MICG3	Standard clothing issued to military personnel (n.d.)
68	P1GASG3	Gasoline and other motor fuel
69	P1LUBG3	Lubricants
70	P1OILG3	Fuel oil
71	P1LPGG3	Liquified petroleum gas and other fuel
72	P1TOBG3	Tobacco products
73	P1SOAG3	Soap
74	P1CSMG3	Cosmetics and perfumes
75	P1OPHG3	Other personal hygiene goods
76	P1SDHG3	Semidurable house furnishings
77	P1CLEG3	Cleaning preparations
78	P1LIGG3	Lighting supplies
79	P1PAPG3	Paper products
80	P1RXDG3	Prescription drugs
81	P1NRXG3	Nonprescription drugs
82	P1MDSG3	Medical supplies
83	P1GYNG3	Gynecological goods
84	P1DOLG3	Toys, dolls, and games
85	P1AMMG3	Sport supplies, including ammunition
86	P1FLMG3	Film and photo supplies
87	P1STSG3	Stationery and school supplies
88	P1GREG3	Greeting cards
89	P1MGZG3	Magazines and sheet music
90	P1NWPG3	Newspapers
91	P1FLOG3	Flowers, seeds, and potted plants

C

Appendix: Social value of uncertainty

C.1 Model definitions

Normalizing constant

The normalizing constant, by definition, is the unconditional probability of $\{y, x_i\}$:

$$f(\mathcal{I}_i) = \int_0^\infty \int_0^\infty \int_{-\infty}^\infty f_{X_i|y,\alpha,\beta,\theta}(x_i) f_{Y|\alpha,\beta,\theta}(y) f_{\Theta|\alpha,\beta}(\theta) f_A(\alpha) f_B(\beta) d\theta d\alpha d\beta \quad \text{(C.1)}$$

Integrating out Θ:

$$f(\mathcal{I}_i) = \int_0^\infty \int_0^\infty f_{X_i|y,\alpha,\beta}(x_i) f_{Y|\alpha,\beta}(y) f_A(\alpha) f_B(\beta) d\alpha d\beta \quad \text{(C.2)}$$

Note that $f_{Y|\alpha,\beta}(y)$ is an improper uniform distribution over $\mathbb{R}$, so the integral is:

$$f(\mathcal{I}_i) = \int_0^\infty \int_0^\infty f_{X_i|y,\alpha,\beta}(x_i) f_A(\alpha) f_B(\beta) d\alpha d\beta \quad \text{(C.3)}$$

Note that $f_{X_i|y,\alpha,\beta}(x_i) = \phi(y - x_i|0, \alpha^{-1} + \beta^{-1})$:

$$f(\mathcal{I}_i) = \int_0^\infty \int_0^\infty \phi(y - x_i|0, \alpha^{-1} + \beta^{-1}) f_A(\alpha) f_B(\beta) d\alpha d\beta \quad \text{(C.4)}$$

which is the desired result.

C.2 Equilibrium existence

In this section, I show that the equilibrium condition in (4.22) satisfies the conditions given in Sloan (1981) that imply the existence of a continuous solution and the convergence of the Nystrom quadrature method.[1] I require an intermediate result, given in the following lemma. As Θ is uniformly distributed over $\mathbb{R}$, I normalize $y = 0$ without loss of generality.

Lemma C.1 *Suppose the model has the information structure given in (4.15) - (4.26), and normalize $y = 0$. Then*

$$\lim_{x_i \to \infty} E_i(\theta) \frac{f_{X_i}(x_i)}{x_i} = 0 \tag{C.5}$$

Proof: The agent's posterior expectation is the expected precision-weighted average of x_i and $y = 0$:

$$E_i(\theta) = x_i \int_0^\infty \int_0^\infty \frac{\beta}{\alpha + \beta} f_{A,B|x_i}(\alpha, \beta) d\alpha d\beta \tag{C.6}$$

Rearranging:

$$E_i(\theta) \frac{f_{X_i}(x_i)}{x_i} = \int_0^\infty \int_0^\infty \frac{\beta}{\alpha + \beta} f_{X_i}(x_i) f_{A,B|x_i}(\alpha, \beta) d\alpha d\beta \tag{C.7}$$

To show the convergence in (C.5), it is sufficient to show convergence of the portion of (C.7) that depends on x_i, which is to show:

$$\lim_{x_i \to \infty} f_{X_i}(x_i) f_{A,B|x_i}(\alpha, \beta) = 0, \ \forall \ \alpha, \beta \in \mathbb{R}^+ \tag{C.8}$$

The limit can be simplified:

$$\lim_{x_i \to \infty} f_{X_i}(x_i) f_{A,B|x_i}(\alpha, \beta) = \lim_{x_i \to \infty} f_{X_i}(x_i) \frac{f_{X_i|\alpha,\beta}(\alpha, \beta) f_A(\alpha) f_B(\beta)}{f_{X_i}(x_i)} \tag{C.9}$$

$$= \lim_{x_i \to \infty} f_{X_i|\alpha,\beta}(\alpha, \beta) f_A(\alpha) f_B(\beta) \tag{C.10}$$

$$= 0 \tag{C.11}$$

[1] The conditions in Sloan (1981) are his equations (2.1), (2.2) and (2.3).

where the last equality comes from the fact that with the normalization $y = 0$, $\lim_{x_i \to \infty} f_{X_i|\alpha,\beta}(\alpha, \beta) = \lim_{x_i \to \infty} \phi(x_i|0, \alpha^{-1} + \beta^{-1}) = 0$. ∎

Lemma C.2 *Suppose the model has the information structure given in (4.15) - (4.26), and normalize $y = 0$. Then there exists an action a_i as a function of x_i in the space of continuous functions.*

Proof: Define a rescaling term as:

$$c_i \equiv \frac{f_{X_i}(x_i)}{x_i^2 + 1} \tag{C.12}$$

Then rewrite the problem as:

$$c_i a_i = r c_i \mathrm{E}_i(\theta) + \frac{(1-r)}{x_i^2 + 1} \int_{-\infty}^{\infty} f_{X_i}(x_i) a_j dF_{x_j|\mathcal{I}_i}(x_j) \tag{C.13}$$

To show the existence of a solution, the following conditions in Sloan (1981), his equations (2.1) - (2.3), are required:

1. $c_i \mathrm{E}_i(\theta)$ and $\frac{(1-r)}{x_i^2+1}$ are continuous and bounded.
2. $\int_{-\infty}^{\infty} f_{X_i}(x_i) f_{x_j|\mathcal{I}_i}(x_j) dx_j < \infty$ for all $x_i \in \mathbb{R}$
3. $\lim_{x_{i'} \to x_i} \int_{-\infty}^{\infty} f_{X_i}(x_i) |f_{x_j|\mathcal{I}_i}(x_j) - f_{x_j|\mathcal{I}_i'}(x_j)| dx_j = 0$ for all $x_i \in \mathbb{R}$
4. $\lim_{x_{i'} \to x_i} \sup_{x_{i'} \geq x_i} \int_{-\infty}^{\infty} f_{X_i}(x_i) |f_{x_j|\mathcal{I}_i}(x_j) - f_{x_j|\mathcal{I}_i'}(x_j)| dx_j = 0$ for all $x_i \in \mathbb{R}$

Condition (1) follows from Lemma C.1.

Condition (2) is satisfied because $f_{X_i}(x_i)$ is strictly positive for all $x \in \mathbb{R}$, is unimodal with a maximum at $x_i = 0$, and is thus bounded:

$$\int_{-\infty}^{\infty} f_{X_i}(x_i) f_{x_j|\mathcal{I}_i}(x_j) dx_j = f_{X_i}(x_i) < \infty \tag{C.14}$$

To show Conditions (3-4), first simplify the integral in (C.13):

$$\int_{-\infty}^{\infty} f_{X_i}(x_i)|f_{x_j|\mathcal{I}_i}(x_j) - f_{x_j|\mathcal{I}_{i'}}(x_j)|dx_j \tag{C.15}$$

$$= \int_{-\infty}^{\infty} \left| \int_{\mathbb{R}\times\mathbb{R}_+^2} \phi(x_j|\theta,\beta^{-1})\phi(0|\theta,\alpha^{-1}) f_A(\alpha) f_B(\beta) \left(\phi(x_i|\theta,\beta^{-1}) - \phi(x_i'|\theta,\beta^{-1}) \frac{f_{X_i}(x_i)}{f_{X_i}(x_i')} \right) d\beta d\alpha d\theta \right| dx_j \tag{C.16}$$

To prove Condition (3), it is sufficient to check the continuity of each component of (C.16).

By the continuity of the Gaussian distribution,

$$\lim_{x_{i'}\to x_i} (\phi(x_i|\theta,\beta^{-1}) - \phi(x_i'|\theta,\beta^{-1})) = 0 \tag{C.17}$$

Also by the continuity of the Gaussian distribution,

$$\lim_{x_{i'}\to x_i} f_{X_i}(x_i') = \lim_{x_{i'}\to x_i} \int_0^\infty \int_0^\infty \phi(x_i'|0,\alpha^{-1}+\beta^{-1}) f_A(\alpha) f_B(\beta) d\alpha d\beta = f_{X_i}(x_i) \tag{C.18}$$

By the continuous mapping theorem, (C.16), (C.17) and (C.18) imply Condition (3).

To prove Condition (4), it is sufficient to check that (C.16) has a bounded derivative at every point in $\mathbb{R}$.

First note that

$$\lim_{x_{i'}\to x_i} \sup_{x_{i'}\geq x_i} (\phi(x_i|\theta,\beta^{-1}) - \phi(x_i'|\theta,\beta^{-1})) = 0 \tag{C.19}$$

Also

$$\lim_{x_{i'}\to x_i} \sup_{x_{i'}\geq x_i} (\phi(x_i'|0,\alpha^{-1}+\beta^{-1}))^{-1} = (\phi(x_i|0,\alpha^{-1}+\beta^{-1}))^{-1} \tag{C.20}$$

Which implies that

$$\lim_{x_{i'}\to x_i} \sup_{x_{i'}\geq x_i} \left(\int_0^\infty \int_0^\infty \phi(x_i'|0,\alpha^{-1}+\beta^{-1}) f_A(\alpha) f_B(\beta) d\alpha d\beta \right)^{-1} \tag{C.21}$$

$$= \left(\int_0^\infty \int_0^\infty \phi(x_i|0,\alpha^{-1}+\beta^{-1}) f_A(\alpha) f_B(\beta) d\alpha d\beta \right)^{-1} \tag{C.22}$$

Hence

$$\lim_{x_{i'}\to x_i} \sup_{x_{i'}\geq x_i} \frac{f_{X_i}(x_i)}{f_{X_i}(x_i')} = \lim_{x_{i'}\to x_i} \sup_{x_{i'}\geq x_i} \frac{\int_0^\infty \int_0^\infty \phi(x_i'|0,\alpha^{-1}+\beta^{-1}) f_A(\alpha) f_B(\beta) d\alpha d\beta}{\int_0^\infty \int_0^\infty \phi(x_i'|0,\alpha^{-1}+\beta^{-1}) f_A(\alpha) f_B(\beta) d\alpha d\beta} = 1 \tag{C.23}$$

Combining (C.16), (C.19) and (C.23) imply Condition (4). ∎

C.3 Model properties

C.3.1 Differentiability assumptions

The welfare results require the assumption that the following derivatives exist:

$$\frac{\partial a_i}{\partial \mu_\alpha}, \frac{\partial a_i}{\partial \mu_\beta}, \frac{\partial a_i}{\partial \tau_\alpha}, \frac{\partial a_i}{\partial \tau_\alpha}, \frac{\partial^2 a_i}{\partial r \partial \mu_\beta}, \frac{\partial^2 a_i}{\partial r \partial \tau_\beta}, \tag{C.24}$$

for $x_i \in \mathbb{R}$, $\mu_\alpha, \mu_\beta, \tau_\alpha, \tau_\beta \in \mathbb{R}^+$ and $r \in [0, 1)$.

C.3.2 Model properties

In this section, I prove basic properties of the model that are required for the welfare results.

Lemma C.3 *Suppose the model has the information structure given in (4.15) - (4.26), and normalize $y = 0$. Further suppose that α^{-1} follows an inverse Gamma distribution as in (4.23), and β^{-1} is fixed. Then, for fixed $i \in [0, 1]$,*

$$\frac{\partial}{\partial \mu_\alpha} E_i \left(\frac{1}{\alpha + \beta} \right) > 0 \tag{C.25}$$

Proof: The proof is in two steps. The first step proves the existence of the derivative in (C.25), and the second step proves the inequality.

Existence

Fix $i \in [0, 1]$. The derivative in (C.25) can be written

$$\frac{\partial}{\partial \mu_\alpha} \left(\int_{-\infty}^{\infty} \left(\frac{1}{\alpha + \beta} \right) f_{A|\mathcal{I}_i}(\alpha) d\alpha \right) \tag{C.26}$$

$$= \frac{\partial}{\partial \mu_\alpha} \left(\int_{-\infty}^{\infty} \left(\frac{1}{\alpha + \beta} \right) f_{X_i|\alpha}(x_i) \frac{f(\alpha)}{f_{X_i}(x_i)} d\alpha \right) \tag{C.27}$$

Note that μ_α only enters into $f_A(\alpha)$. So it is sufficient to show that $f_A(\alpha)$ is differentiable by μ_α. This is the case because $f_A(\alpha)$ is differentiable by $k_\alpha > 2$ and $\ell_\alpha > 0$ (see (4.23)), and because $k_\alpha > 2$ and $\ell_\alpha > 0$ are differentiable by μ_α (see (4.26)).

Inequality

Write the inverse Gamma pdf from (4.23) in terms of k_α and μ_α:

$$f_{A^{-1}}(\alpha^{-1}) = \frac{(\mu_\alpha(k_\alpha - 1))^{k_\alpha}}{\Gamma(k_\alpha)} x^{-k_\alpha - 1} \exp\left(-\frac{\mu_\alpha(k_\alpha - 1)}{x} \right) \tag{C.28}$$

Consider the derivative:

$$\frac{\partial f_{A^{-1}}(\alpha^{-1})}{\partial \mu_\alpha} = x^{-k_\alpha - 1} \frac{k_\alpha \mu_\alpha^{k_\alpha - 1}(k_\alpha - 1)^{k_\alpha}}{\Gamma(k_\alpha)} \exp\left(-\frac{\mu_\alpha(k_\alpha - 1)}{x}\right) \tag{C.29}$$

$$- x^{-k_\alpha - 1} \frac{(\mu_\alpha(k_\alpha - 1))^{k_\alpha}}{\Gamma(k_\alpha)} \exp\left(-\frac{\mu_\alpha(k_\alpha - 1)}{x}\right) \frac{k_\alpha - 1}{x} \tag{C.30}$$

It is the case that

$$\frac{\partial f_{A^{-1}}(\alpha^{-1})}{\partial \mu_\alpha} < 0, \ \alpha^{-1} < \mu_\alpha(\kappa_\alpha - 1) \tag{C.31}$$

$$\frac{\partial f_{A^{-1}}(\alpha^{-1})}{\partial \mu_\alpha} > 0, \ \alpha^{-1} > \mu_\alpha(\kappa_\alpha - 1) \tag{C.32}$$

The density always integrates to 1, and so

$$\frac{\partial}{\partial \mu_\alpha} \int_0^\infty \left(\frac{f_{X_i|\alpha^{-1}}(x_i)}{f_{X_i}(x_i)} f_{A^{-1}}(\alpha^{-1}) \right) d\alpha = 0 \tag{C.33}$$

Combining (C.31) - (C.33) yields

$$- \frac{\partial}{\partial \mu_\alpha} \int_0^{\mu_\alpha(\kappa_\alpha - 1)} \left(\frac{f_{X_i|\alpha^{-1}}(x_i)}{f_{X_i}(x_i)} f_{A^{-1}}(\alpha^{-1}) \right) d\alpha \tag{C.34}$$

$$= \frac{\partial}{\partial \mu_\alpha} \int_{\mu_\alpha(\kappa_\alpha - 1)}^\infty \left(\frac{f_{X_i|\alpha^{-1}}(x_i)}{f_{X_i}(x_i)} f_{A^{-1}}(\alpha^{-1}) \right) d\alpha \tag{C.35}$$

Note that $\frac{1}{\alpha+\beta}$ is increasing in α^{-1}. So,

$$- \frac{\partial}{\partial \mu_\alpha} \int_0^{\mu_\alpha(\kappa_\alpha - 1)} \left(\frac{1}{\alpha+\beta} \frac{f_{X_i|\alpha^{-1}}(x_i)}{f_{X_i}(x_i)} f_{A^{-1}}(\alpha^{-1}) \right) d\alpha \tag{C.36}$$

$$< \frac{\partial}{\partial \mu_\alpha} \int_{\mu_\alpha(\kappa_\alpha - 1)}^\infty \left(\frac{1}{\alpha+\beta} \frac{f_{X_i|\alpha^{-1}}(x_i)}{f_{X_i}(x_i)} f_{A^{-1}}(\alpha^{-1}) \right) d\alpha \tag{C.37}$$

where both terms are positive. Rearranging,

$$\frac{\partial}{\partial \mu_\alpha} \mathrm{E}_i \left(\frac{1}{\alpha+\beta} \right) > 0 \tag{C.38}$$

which is the desired result. ∎

Lemma C.4 *Suppose the model has the information structure given in (4.15) - (4.26), and normalize $y = 0$. Further suppose that β^{-1} follows an inverse Gamma distribution as in (4.23), and α^{-1} is fixed. Then, for fixed $i \in [0, 1]$,*

$$\frac{\partial}{\partial \mu_\beta} E_i \left(\frac{1}{\alpha+\beta} \right) > 0 \tag{C.39}$$

Proof: Because α and β enter symmetrically into $\mathrm{E}_i(\cdot)$, the logic is analogous to that of Lemma C.3. ∎

Lemma C.5 *Suppose the model has the information structure given in (4.15) - (4.26), and normalize $y = 0$. Further suppose that α^{-1} follows an inverse Gamma distribution as in (4.23), and β^{-1} is fixed. Then*

$$\frac{\partial}{\partial \tau_\alpha} E_i\left(\frac{\alpha}{\alpha+\beta}\right) > 0 \tag{C.40}$$

Proof: The proof is in two steps. The first step proves the existence of the derivative in (C.40). The second step proves the inequality in (C.40).

Existence

Fix $i \in [0,1]$. The derivative in (C.40) can be written

$$\frac{\partial}{\partial \tau_\alpha}\left(\int_{-\infty}^{\infty}\left(\frac{\alpha}{\alpha+\beta}\right) f_{A|\mathcal{I}_i}(\alpha)d\alpha\right) \tag{C.41}$$

$$= \frac{\partial}{\partial \tau_\alpha}\left(\int_{-\infty}^{\infty}\left(\frac{\alpha}{\alpha+\beta}\right) f_{X_i|\alpha}(x_i)\frac{f_A(\alpha)}{f_{X_i}(x_i)}d\alpha\right) \tag{C.42}$$

Note that $f_{X_i|\alpha}(x_i)$ and $f_{X_i}(x_i) = \int_0^\infty f_{X_i|\alpha}(x_i)f_A(\alpha)d\alpha$ are differentiable by x_i, as this is a property of the Gaussian distribution. Also, $f_A(\alpha)$ is differentiable by τ_α, as this is a property of the inverse Gamma distribution. Together, this proves the existence of the derivative in (C.40).

Inequality

Define the following function $g(\alpha^{-1})$, and note that it is composed of terms dependent on α^{-1}, and also a positive function $\gamma(x_i)$ that does not depend on α^{-1}:

$$g(\alpha^{-1}) \equiv \left(\frac{\alpha}{\alpha+\beta}\right)\frac{f_{X_i|\alpha}(x_i)}{f_{X_i}(x_i)} = \frac{(\alpha^{-1})^{-0.5}\exp((\alpha^{-1})^{-1})}{(\alpha^{-1})^{-1}+\beta}\gamma(x_i) \tag{C.43}$$

Examine the first and second derivatives:

$$\frac{\partial g(\alpha^{-1})}{\partial \alpha^{-1}} = -\gamma(x_i)\frac{\exp(\alpha)\left(3\alpha^{-2}\beta + 2\alpha^{-1}\beta + \alpha^{-1} + 2\right)}{2\alpha^{-5/2}(\alpha\beta+1)^2} \tag{C.44}$$

$$\frac{\partial^2 g(\alpha^{-1})}{\partial(\alpha^{-1})^2} = \gamma(x_i)\frac{\exp(\alpha)\left(15\alpha^{-4}\beta^2 + 10\alpha^{-3}\beta(2\beta+1) + \alpha^{-2}\left(4\beta^2+32\beta+3\right) + 4\alpha^{-1}(2\beta+3) + 4\right)}{4\alpha^{-9/2}(\alpha^{-1}\beta+1)^3} \tag{C.45}$$

Note that $g(\alpha^{-1})$ is a nonincreasing and convex function of α^{-1}.

Suppose that $\tau_\alpha < \tau'_\alpha$. Then $f_{A^{-1}|\tau'_\alpha}(\alpha^{-1})$ is a mean-preserving spread of $f_{A^{-1}|\tau_\alpha}(\alpha^{-1})$. Thus $f_{A^{-1}|\tau_\alpha}(\alpha^{-1})$ second-order stochastically dominates $f_{A^{-1}|\tau'_\alpha}(\alpha^{-1})$. Because $g(\alpha^{-1})$ is nonincreasing and convex, then

$$\int_0^\infty g(\alpha^{-1})f_{A^{-1}|\tau_\alpha}(\alpha^{-1}) > \int_0^\infty g(\alpha^{-1})f_{A^{-1}|\tau'_\alpha}(\alpha^{-1}) \tag{C.46}$$

Substituting in the definition of $g(\alpha^{-1})$ yields

$$\frac{\partial}{\partial \tau_\alpha} \mathrm{E}_i \left(\frac{\alpha}{\alpha + \beta} \right) > 0 \tag{C.47}$$

which is the desired result. ∎

Lemma C.6 *Suppose the model has the information structure given in (4.15 - 4.26), and normalize $y = 0$. Further suppose that α^{-1} is fixed and β^{-1} follows an inverse Gamma distribution as in (4.23). Then, for fixed $i \in [0, 1]$,*

$$\frac{\partial}{\partial \tau_\beta} E_i \left(\frac{\beta}{\alpha + \beta} \right) > 0 \tag{C.48}$$

Proof: This is an analogue of Lemma C.5. ∎

C.4 Welfare results

Lemma C.7 ***Effect of public signal mean-uncertainty on accuracy*** *Suppose the model has the information structure given in (4.15) - (4.26). Further suppose that β^{-1} is fixed, and α^{-1} follows an inverse Gamma distribution as specified in (4.23). Then*

$$\frac{\partial E[(E_i(\theta) - \theta)^2]}{\partial \mu_\alpha} > 0 \tag{C.49}$$

Proof: Fix $i \in [0, 1]$. Note that

$$\begin{aligned}
\mathrm{E}_i((\mathrm{E}_i(\theta) - \theta)^2) &= \mathrm{E}_i \left(\left(\frac{\alpha y + \beta x}{\alpha + \beta} - \theta \right)^2 \right) && \text{(C.50)} \\
&= \mathrm{E}_i \left(\left(\frac{\alpha(y - \theta) + \beta(x - \theta)}{\alpha + \beta} \right)^2 \right) && \text{(C.51)} \\
&= \mathrm{E}_i \left(\frac{\alpha^2 \mathrm{Var}(\eta) + \beta^2 \mathrm{Var}(\epsilon_i)}{(\alpha + \beta)^2} \right) && \text{(C.52)} \\
&= \mathrm{E}_i \left(\frac{\alpha^2 \alpha^{-1} + \beta^2 \beta^{-1}}{(\alpha + \beta)^2} \right) && \text{(C.53)} \\
&= \mathrm{E}_i \left(\frac{1}{\alpha + \beta} \right) && \text{(C.54)}
\end{aligned}$$

By Lemma C.3,

$$\frac{\partial}{\partial \mu_\alpha} \mathrm{E}_i \left(\frac{1}{\alpha + \beta} \right) > 0 \tag{C.55}$$

Combining (C.54) and (C.55) gives

$$\frac{\partial}{\partial \mu_\alpha} \mathrm{E}_i((\mathrm{E}_i(\theta) - \theta)^2) > 0 \tag{C.56}$$

As we have conditioned on i, which is invariant to μ_α, then we can take expectations over i and the inequality still holds:

$$\frac{\partial \mathrm{E}[\mathrm{E}_i((\mathrm{E}_i(\theta) - \theta)^2)]}{\partial \mu_\alpha} > 0 \tag{C.57}$$

Iterating expectations yields

$$\frac{\partial \mathrm{E}[(\mathrm{E}_i(\theta) - \theta)^2]}{\partial \mu_\alpha} > 0 \tag{C.58}$$

which is the desired result. ∎

Lemma C.8 *Suppose the model has the information structure given in (4.15) - (4.26). Further suppose that α^{-1} is fixed, and β^{-1} follows an inverse Gamma distribution as specified in (4.23). Then*

$$\frac{\partial E[(E_i(\theta) - a_i)^2]}{\partial \mu_\beta \partial r} \geq 0 \tag{C.59}$$

Proof: This proof is in two steps. The first is an intermediate monotonicity result. The second step proves the inequality in (C.59). Without loss of generality, normalize $y = 0$.

Monotonicity

Fix $x_i > 0$. From Lemma C.4 it is the case that

$$\frac{\partial}{\partial \mu_\beta} \mathrm{E}_i(\theta) \Big|_{x_i} = \frac{\partial}{\partial \mu_\beta} \mathrm{E}_i\left(\frac{\beta x_i}{\alpha + \beta}\right) \Big|_{x_i} < 0 \tag{C.60}$$

Recall from the definition of the utility function in (4.1) that $\mathrm{E}(u_i)$ is supermodular in $(a_i, \mathrm{E}_i(\theta))$, in the sense that

$$\frac{\partial^2 \mathrm{E}(u_i)}{\partial a_i \partial \mathrm{E}_i(\theta)} \Big|_{x_i} > 0 \tag{C.61}$$

Combining (C.60) and (C.61) yields

$$\frac{\partial^2 \mathrm{E}(u_i)}{\partial a_i \partial \mu_\beta} \Big|_{x_i} < 0 \tag{C.62}$$

Consequently, for any fixed $x_i > 0$, it will be the case that the equilibrium behavior a_i will be monotone in μ_β:

$$\frac{\partial a_i}{\partial \mu_\beta} \Big|_{x_i} < 0 \tag{C.63}$$

Inequality

Now differentiate (4.22), noting that $\partial \mathrm{E}_i(\theta)/\partial r = \partial f_{x_j|\mathcal{I}_i}(x_j)/\partial r = 0$:

$$\frac{\partial a_i}{\partial r} = \mathrm{E}_i(a_j) + r \int_{\mathbb{R}} \frac{\partial a_j}{\partial r} f_{x_j|\mathcal{I}_i}(x_j) dx_j \tag{C.64}$$

This is an inhomogenous Fredholm equation of the second kind. It is isomorphic to the equilibrium condition in (4.22), where κ_i is replaced by $\partial \kappa_i/\partial r$, and $(1 - r)\mathrm{E}_i(\theta)$ is replaced by $\mathrm{E}_i(a_j)$.

As a result, we can apply the same logic used in the first step of this proof to the game described by (C.64). The analogous result to (C.63) is the result that

$$\left.\frac{\partial^2 E_i(a_j)}{\partial r \partial \mu_\beta}\right|_{x_i} < 0 \tag{C.65}$$

Now note that, because $E_i(\theta)$ does not contain the strategic component of the model, and is unaffected by r,

$$\left.\frac{\partial^2 E_i(\theta)}{\partial r \partial \mu_\beta}\right|_{x_i} = 0 \tag{C.66}$$

This implies that, for some $x_i > 0$, since $E_i(a_j) < E_i(\theta)$,

$$\left.\frac{\partial (\mathrm{E}_i(\theta) - a_i)^2}{\partial \mu_\beta \partial r}\right|_{x_i} \geq 0 \tag{C.67}$$

The model is symmetric about 0, and so the analogous argument holds for $x_i \leq 0$.

Now note that, because the distribution of x_i is invariant to r, that

$$\left.\frac{\partial f_{X_i}(x_i)}{\partial \mu_\beta \partial r}\right|_{x_i} = 0 \tag{C.68}$$

Combining (C.67) and (C.68) implies the inequality is true in expectation over X_i:

$$\frac{\partial \mathrm{E}[(\mathrm{E}_i(\theta) - a_i)^2]}{\partial \mu_\beta \partial r} \geq 0 \tag{C.69}$$

which is the desired result. ∎

Lemma C.9 ***Effect of private signal mean-uncertainty on distortion*** *Suppose the model is defined as in (4.15) to (4.22). Further suppose that α^{-1} is fixed, and β^{-1} follows an inverse Gamma distribution as specified in (4.23). Then*

$$\frac{\partial E[(a_i - E_i(\theta))^2]}{\partial \mu_\beta} \geq 0 \tag{C.70}$$

Proof: From Lemma C.8,

$$\frac{\partial \mathrm{E}[(\mathrm{E}_i(\theta) - a_i)^2]}{\partial \mu_\beta \partial r} \geq 0 \tag{C.71}$$

Therefore, to show the desired result, it is sufficient to show that

$$\frac{\partial \mathrm{E}[\mathrm{E}_i(\theta) - a_i)^2]}{\partial \mu_\beta} \geq 0 \text{ at } r = 0 \tag{C.72}$$

When $r = 0$, then the beauty-contest term is 0, and $\mathrm{E}_i(\theta) = a_i$, so

$$\frac{\partial \mathrm{E}[(\mathrm{E}_i(\theta) - a_i)^2]}{\partial \mu_\beta} = 0 \text{ at } r = 0 \tag{C.73}$$

Hence

$$\frac{\partial \mathrm{E}[(\mathrm{E}_i(\theta) - a_i)^2]}{\partial \mu_\beta} \geq 0 \tag{C.74}$$

which proves the desired result. ∎

Lemma C.10 ***Effect of private signal mean-uncertainty on accuracy*** *Suppose the model is defined as in (4.15) to (4.22). Further suppose that α^{-1} is fixed, and β^{-1} follows an inverse Gamma distribution as specified in (4.23). Then*

$$\frac{\partial E[(E_i(\theta) - \theta)^2]}{\partial \mu_\beta} > 0 \tag{C.75}$$

Proof: This has the identical logic to Lemma C.7, because x_i and y enter symmetrically into the function $\mathrm{E}_i(\theta)$. ∎

Lemma C.11 ***Effect of public signal variance-uncertainty on accuracy*** *Suppose the model is defined as in (4.15) to (4.22). Further suppose that β^{-1} is fixed, and α^{-1} follows an inverse Gamma distribution as specified in (4.23). Then*

$$\frac{\partial E[(E_i(\theta) - \theta)^2]}{\partial \tau_\alpha} < 0 \tag{C.76}$$

Proof: From Lemma C.5 we have that

$$\frac{\partial}{\partial \tau_\alpha} \mathrm{E}_i \left(\frac{\beta}{\alpha + \beta} \right) > 0 \tag{C.77}$$

Note that

$$\frac{\partial}{\partial \tau_\alpha} \mathrm{Var}\,(x_i) = 0 \tag{C.78}$$

Thus

$$\frac{\partial}{\partial \tau_\alpha} \text{Var}\left(x_i \text{E}_i\left(\frac{\beta}{\alpha+\beta}\right)\right) > 0 \tag{C.79}$$

Which is to say that

$$\frac{\partial}{\partial \tau_\alpha} \text{Var}\left(\text{E}_i(\theta)\right) > 0 \tag{C.80}$$

Using the law of total variation, $\text{Var}(\theta) = \text{E}(\text{Var}(\theta|x_i)) + \text{Var}(\text{E}(\theta|x_i))$, we have

$$0 = \frac{\partial \text{Var}(\theta)}{\partial \tau_\alpha} = \frac{\partial \text{E}(\text{Var}(\theta|x_i))}{\partial \tau_\alpha} + \frac{\partial \text{Var}(\text{E}(\theta|x_i))}{\partial \tau_\alpha} \tag{C.81}$$

This implies that

$$\frac{\partial \text{E}(\text{Var}(\theta|x_i))}{\partial \tau_\alpha} < 0 \tag{C.82}$$

which is the desired result. ∎

Lemma C.12 *Suppose the model is defined as in (4.15) to (4.22). Further suppose that α^{-1} is fixed, and β^{-1} follows an inverse Gamma distribution as specified in (4.23). Then*

$$\frac{\partial E[(E_i(\theta) - a_i)^2]}{\partial \tau_\beta \partial r} < 0 \tag{C.83}$$

Proof: This proof is in two steps. The first is an intermediate monotonicity result. The second step proves the inequality in (C.83). Without loss of generality, normalize $y = 0$.

Monotonicity

Fix $x_i > 0$. From Lemma C.6 it is the case that

$$\frac{\partial}{\partial \tau_\beta} \text{E}_i\left(\theta\right)\Big|_{x_i} = \frac{\partial}{\partial \tau_\beta} \text{E}_i\left(\frac{\beta x_i}{\alpha+\beta}\right)\Big|_{x_i} > 0 \tag{C.84}$$

Recall from the definition of the utility function in (4.1) that $\text{E}(u_i)$ is supermodular in $(a_i, \text{E}_i(\theta))$, in the sense that

$$\frac{\partial^2 \text{E}(u_i)}{\partial a_i \partial \text{E}_i(\theta)}\Big|_{x_i} > 0 \tag{C.85}$$

Combining (C.84) and (C.85) yields

$$\frac{\partial^2 \text{E}(u_i)}{\partial a_i \partial \tau_\beta}\Big|_{x_i} > 0 \tag{C.86}$$

Consequently, for any fixed $x_i > 0$, it will be the case that the equilibrium behavior a_i will be monotone in μ_β:

$$\left.\frac{\partial a_i}{\partial \tau_\beta}\right|_{x_i} > 0 \tag{C.87}$$

Inequality

Now differentiate (4.22), noting that $\partial \mathrm{E}_i(\theta)/\partial r = \partial f_{x_j|\mathcal{I}_i}(x_j)/\partial r = 0$:

$$\frac{\partial a_i}{\partial r} = \mathrm{E}_i(a_j) + r\int_{\mathbb{R}} \frac{\partial a_j}{\partial r} f_{x_j|\mathcal{I}_i}(x_j)dx_j \tag{C.88}$$

This is an inhomogenous Fredholm equation of the second kind. It is isomorphic to the equilibrium condition in (4.22), where κ_i is replaced by $\partial\kappa_i/\partial r$, and $(1-r)\mathrm{E}_i(\theta)$ is replaced by $\mathrm{E}_i(a_j)$.

As a result, we can apply the same logic used in the first step of this proof to the game described by (C.88). The analogous result to (C.87) is the result that

$$\left.\frac{\partial^2 E_i(a_j)}{\partial r \partial \tau_\beta}\right|_{x_i} > 0 \tag{C.89}$$

Now note that, because $E_i(\theta)$ does not contain the strategic component of the model, and is unaffected by r,

$$\left.\frac{\partial^2 E_i(\theta)}{\partial r \partial \tau_\beta}\right|_{x_i} = 0 \tag{C.90}$$

This implies that, for some $x_i > 0$, since $E_i(a_j) < E_i(\theta)$,

$$\left.\frac{\partial (\mathrm{E}_i(\theta) - a_i)^2}{\partial \mu_\beta \partial r}\right|_{x_i} < 0 \tag{C.91}$$

The model is symmetric about 0, and so the analogous argument holds for $x_i \leq 0$.

Now note that, because the distribution of x_i is invariant to r, that

$$\left.\frac{\partial f_{X_i}(x_i)}{\partial \tau_\beta \partial r}\right|_{x_i} = 0 \tag{C.92}$$

Combining (C.91) and (C.92) implies the inequality is true in expectation over X_i:

$$\frac{\partial \mathrm{E}[(\mathrm{E}_i(\theta) - a_i)^2]}{\partial \tau_\beta \partial r} < 0 \tag{C.93}$$

which is the desired result. ∎

Lemma C.13 ***Effect of private signal mean-uncertainty on distortion*** *Suppose the model is defined as in (4.15) to (4.22). Further suppose that α^{-1} is fixed, and β^{-1} follows an inverse Gamma distribution as specified in (4.23). Then*

$$\frac{\partial E[(a_i - E_i(\theta))^2]}{\partial \tau_\beta} \leq 0 \tag{C.94}$$

Proof: From Lemma C.12,

$$\frac{\partial \mathrm{E}[(\mathrm{E}_i(\theta) - a_i)^2]}{\partial \tau_\beta \partial r} < 0 \tag{C.95}$$

Therefore, it is sufficient to show that

$$\frac{\partial \mathrm{E}[(\mathrm{E}_i(\theta) - a_i)^2]}{\partial \tau_\beta} < 0 \text{ at } r = 0 \tag{C.96}$$

When $r = 0$, then the weight on the beauty-contest term is 0, and $\mathrm{E}_i(\theta) = a_i$, so

$$\frac{\partial \mathrm{E}[(\mathrm{E}_i(\theta) - a_i)^2]}{\partial \tau_\beta} = 0 \text{ at } r = 0 \tag{C.97}$$

Hence

$$\frac{\partial \mathrm{E}[(\mathrm{E}_i(\theta) - a_i)^2]}{\partial \tau_\beta} \leq 0 \tag{C.98}$$

which is the desired result. ∎

Lemma C.14 *Given the model as described in (4.15) - (4.22), a prior $f_B(\beta)$ that satisfies the prior conditions in (C.3), and also $\tau_\alpha = 0$ (α is known), then*

$$\frac{\partial E[(E_i(\theta) - \theta)^2]}{\partial \tau_\beta} < 0 \tag{C.99}$$

Proof: This follows the same logic as Lemma (C.11), because x_i and y enter symmetrically into the function $\mathrm{E}_i(\theta)$. ∎

Bibliography

Adam, K. (2007). Optimal monetary policy with imperfect common knowledge. *Journal of monetary Economics*, 54(2):267–301.

Amato, J. D. and Shin, H. S. (2003). Public and private information in monetary policy models. *Mimeo*.

Andrade, P., Crump, R. K., Eusepi, S., and Moench, E. (2013). Noisy information and fundamental disagreement. Technical report, Staff Report, Federal Reserve Bank of New York.

Angeletos, G.-M. and La'O, J. (2013). Sentiments. *Econometrica*, 81(2):739–779.

Angeletos, G.-M. and Pavan, A. (2003). The value of information and coordination in economies with investment complementarities. Technical report, Massachusetts Institute of Technology.

Angeletos, G.-M. and Pavan, A. (2004). Transparency of information and coordination in economies with investment complementarities. *The American Economic Review*, 94(2):91–98.

Angeletos, G.-M. and Pavan, A. (2007). Efficient use of information and social value of information. *Econometrica*, 75(4):1103–1142.

Ashiya, M. (2009). Strategic bias and professional affiliations of macroeconomic forecasters. *Journal of Forecasting*, 28(2):120–130.

Atkeson, A. and Ohanian, L. E. (2001). Are Phillips Curves useful for forecasting inflation? *Federal Reserve Bank of Minneapolis Quarterly Review*, 25(1):2–11.

Bachmann, R. and Bayer, C. (2011). Uncertainty business cycles-really? Technical report, National Bureau of Economic Research.

Bachmann, R., Elstner, S., and Sims, E. R. (2013). Uncertainty and economic activity: Evidence from business survey data. *American Economic Journal: Macroeconomics*, 5(2):217–249.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636.

Basu, S. and Bundick, B. (2017). Uncertainty shocks in a model of effective demand. *Econometrica*, 85(3):937–958.

Beaudry, P. and Portier, F. (2014). News-driven business cycles: insights and challenges. *Journal of Economic Literature*, 52(4):993–1074.

Bekaert, G. and Hoerova, M. (2014). The VIX, the variance premium and stock market volatility. *Journal of Econometrics*, 183(2):181–192.

Bekaert, G., Hoerova, M., and Duca, M. L. (2013). Risk, uncertainty and monetary policy. *Journal of Monetary Economics*, 60(7):771–788.

Berger, D., Dew-Becker, I., and Giglio, S. (2017). Uncertainty shocks as second-moment news shocks. *Mimeo*.

Blinder, A. S., Ehrmann, M., Fratzscher, M., De Haan, J., and Jansen, D.-J. (2008). Central bank communication and monetary policy: A survey of theory and evidence. *Journal of Economic Literature*, 46(4):910–945.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, 77(3):623–685.

Bloom, N., Floetotto, M., Jaimovich, N., Saporta-Eksten, I., and Terry, S. J. (2012). Really uncertain business cycles. Technical report, National Bureau of Economic Research.

Boivin, J., Giannoni, M. P., and Mihov, I. (2009). Sticky prices and monetary policy: Evidence from disaggregated us data. *The American Economic Review*, 99(1):350–384.

Born, B. and Pfeifer, J. (2014). Policy risk and the business cycle. *Journal of Monetary Economics*, 68:68–85.

Capistrán, C. and Timmermann, A. (2009). Disagreement and biases in inflation expectations. *Journal of Money, Credit and Banking*, 41(2-3):365–396.

Caplin, A., Leahy, J., and Matějka, F. (2014). Learning from market share when consumers are rationally inattentive.

Carr, P. and Wu, L. (2008). Variance risk premiums. *The Review of Financial Studies*, 22(3):1311–1341.

Carroll, C. D. (2003). Macroeconomic expectations of households and professional forecasters. *The Quarterly Journal of Economics*, pages 269–298.

Clarida, R., Gali, J., and Gertler, M. (1999). The science of monetary policy: a new Keynesian perspective. Technical report, National bureau of economic research.

Coibion, O. and Gorodnichenko, Y. (2012). What can survey forecasts tell us about information rigidities? *Journal of Political Economy*, 120(1):116–159.

Coibion, O. and Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *The American Economic Review*, 105(8):2644–2678.

Cover, T. M. and Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons.

Crowe, C. (2010). Testing the transparency benefits of inflation targeting: Evidence from private sector forecasts. *Journal of Monetary Economics*, 57(2):226–232.

Cukierman, A. and Wachtel, P. (1979). Differential inflationary expectations and the variability of the rate of inflation: Theory and evidence. *The American Economic Review*, pages 595–609.

Cunha, F. and Heckman, J. J. (2007). The evolution of inequality, heterogeneity and uncertainty in labor earnings in the us economy. Technical report, National Bureau of Economic Research.

Davis, S. J. and Kahn, J. A. (2008). Interpreting the great moderation: Changes in the volatility of economic activity at the macro and micro levels. *The Journal of Economic Perspectives*, 22(4):155–180.

Dovern, J. (2014). A multivariate analysis of forecast disagreement: Confronting models of disagreement with SPF data. *Mimeo*.

Dovern, J., Fritsche, U., and Slacalek, J. (2012). Disagreement among forecasters in G7 countries. *Review of Economics and Statistics*, 94(4):1081–1096.

Drechsler, I. (2013). Uncertainty, time-varying fear, and asset prices. *The Journal of Finance*, 68(5):1843–1889.

Drechsler, I. and Yaron, A. (2010). What's vol got to do with it. *The Review of Financial Studies*, 24(1):1–45.

Ehrbeck, T. and Waldmann, R. (1996). Why are professional forecasters biased? Agency versus behavioral explanations. *The Quarterly Journal of Economics*, pages 21–40.

Eijffinger, S. C. and Geraats, P. M. (2006). How transparent are central banks? *European Journal of Political Economy*, 22(1):1–21.

Elliott, G., Komunjer, I., and Timmermann, A. (2008). Biases in macroeconomic forecasts: irrationality or asymmetric loss? *Journal of the European Economic Association*, 6(1):122–157.

Fernández-Villaverde, J., Guerrón-Quintana, P., Kuester, K., and Rubio-Ramírez, J. (2015). Fiscal volatility shocks and economic activity. *The American Economic Review*, 105(11):3352–3384.

Fernández-Villaverde, J., Guerrón-Quintana, P., Rubio-Ramírez, J. F., and Uribe, M. (2011). Risk matters: The real effects of volatility shocks. *The American Economic Review*, 101(6):2530–2561.

Garthwaite, P. H., Kadane, J. B., and O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, 100(470):680–701.

Harrison, J. and West, M. (1999). *Bayesian forecasting and dynamic models*, volume 1030. Springer.

Hellwig, C. (2005). Heterogeneous information and the welfare effects of public information disclosures. *Mimeo*.

Kahn, J. A. (2008). Durable goods inventories and the great moderation. *Mimeo*.

Kahn, J. A., McConnell, M. M., and Perez-Quiros, G. e. a. (2002). On the causes of the increased stability of the US economy. *Federal Reserve Bank of New York Economic Policy Review*, 8(1):183–202.

Kim, C.-J. and Nelson, C. R. (1999). Has the US economy become more stable? a Bayesian approach based on a Markov-switching model of the business cycle. *Review of Economics and Statistics*, 81(4):608–616.

Lahiri, K. and Sheng, X. (2008). Evolution of forecast disagreement in a Bayesian learning model. *Journal of Econometrics*, 144(2):325–340.

Lorenzoni, G. (2009). A theory of demand shocks. *American Economic Review*, 99(5):2050–84.

Maćkowiak, B., Matejka, F., and Wiederholt, M. (2016). The Rational Inattention filter. *Mimeo*.

Maćkowiak, B., Moench, E., and Wiederholt, M. (2009). Sectoral price data and models of price setting. *Journal of Monetary Economics*, 56:S78–S99.

Maćkowiak, B. and Wiederholt, M. (2009). Optimal sticky prices under Rational Inattention. *The American Economic Review*, 99(3):769–803.

Maćkowiak, B. and Wiederholt, M. (2012). Information processing and limited liability. *The American Economic Review*, 102(3):30–34.

Maćkowiak, B. and Wiederholt, M. (2015). Business cycle dynamics under Rational Inattention. *The Review of Economic Studies*, page rdv027.

Malhotra, D., Loewenstein, G., and O'donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of economic literature*, 40(2):351–401.

Mankiw, N. G. and Reis, R. (2001). Sticky information versus sticky prices: a proposal to replace the New Keynesian Phillips curve. Technical report, National Bureau of Economic Research.

Mankiw, N. G., Reis, R., and Wolfers, J. (2003). Disagreement about inflation expectations. Technical report, National Bureau of Economic Research.

Matějka, F. (2015a). Rationally inattentive seller: Sales and discrete pricing. *The Review of Economic Studies*, page rdv049.

Matějka, F. (2015b). Rigid pricing and rationally inattentive consumer. *Journal of Economic Theory*, 158:656–678.

Matějka, F. and McKay, A. (2012). Simple market equilibria with rationally inattentive consumers. *The American Economic Review*, 102(3):24–29.

Matějka, F. and McKay, A. (2014). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *The American Economic Review*, 105(1):272–298.

Matějka, F., Sims, C. A., et al. (2010). Discrete actions in information-constrained tracking problems. *Princeton University manuscript*.

Matějka, F. and Tabellini, G. (2016). Electoral competition with Rationally Inattentive voters. *Mimeo*.

McConnell, M. M. and Perez-Quiros, G. (2000). Output fluctuations in the united states: What has changed since the early 1980's? *The American Economic Review*, 90(5):1464–1476.

Morris, S. and Shin, H. S. (2002). Social value of public information. *The American Economic Review*, 92(5):1521–1534.

Morris, S., Shin, H. S., and Tong, H. (2006). Social value of public information: Morris and shin (2002) is actually pro-transparency, not con: Reply. *The American Economic Review*, 96(1):453–455.

Nimark, K. (2014). Man-bites-dog business cycles. *The American Economic Review*, 104(8):2320–2367.

Paciello, L. and Wiederholt, M. (2013). Exogenous information, endogenous information and optimal monetary policy. *The Review of Economic Studies*, page rdt024.

Pastor, L. and Veronesi, P. (2012). Uncertainty about government policy and stock prices. *The Journal of Finance*, 67(4):1219–1264.

Pástor, L. and Veronesi, P. (2013). Political uncertainty and risk premia. *Journal of Financial Economics*, 110(3):520–545.

Patton, A. J. and Timmermann, A. (2007). Testing forecast optimality under unknown loss. *Journal of the American Statistical Association*, 102(480).

Patton, A. J. and Timmermann, A. (2010). Why do forecasters disagree? Lessons from the term structure of cross-sectional dispersion. *Journal of Monetary Economics*, 57(7):803–820.

Peng, L. and Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, 80(3):563–602.

Pesaran, M. H. and Weale, M. (2006). Survey expectations. *Handbook of economic forecasting*, 1:715–776.

Pierdzioch, C. and Rülke, J.-C. (2012). Forecasting stock prices: Do forecasters herd? *Economics Letters*, 116(3):326–329.

Pons-Novell, J. (2003). Strategic bias, herding behaviour and economic forecasts. *Journal of Forecasting*, 22(1):67–77.

Popescu, A. and Rafael Smets, F. (2010). Uncertainty, risk-taking, and the business cycle in Germany. *CESifo Economic Studies*, 56(4):596–626.

Reis, R. (2013). Central bank design. *The Journal of Economic Perspectives*, 27(4):17–43.

Roca, M. (2010). Transparency and monetary policy with imperfect common knowledge. Technical Report 10-91.

Sims, C. A. (1998). Stickiness. In *Carnegie-Rochester Conference Series on Public Policy*, volume 49, pages 317–356.

Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics*, 50(3):665–690.

Sims, C. A. (2006). Rational inattention: Beyond the linear-quadratic case. *The American economic review*, 96(2):158–163.

Sims, C. A. et al. (2005). *Rational inattention: a research agenda*. Deutsche Bundesbank.

Sims, C. A. et al. (2010). Rational inattention and monetary economics. *Handbook of Monetary Economics*, 3:155–181.

Sloan, I. H. (1981). Quadrature methods for integral equations of the second kind over infinite intervals. *Mathematics of computation*, 36(154):511–523.

Souleles, N. S. (2004). Expectations, heterogeneous forecast errors, and consumption: Micro evidence from the Michigan consumer sentiment surveys. *Journal of Money, Credit and Banking*, pages 39–72.

Steiner, J., Stewart, C., Matejka, F., et al. (2015). Rational inattention dynamics: Inertia and delay in decision-making. Technical report.

Stock, J. H. and Watson, M. W. (2003). Has the business cycle changed and why? In *NBER Macroeconomics Annual 2002, Volume 17*, pages 159–230. MIT Press.

Stock, J. H. and Watson, M. W. (2007). Why has US inflation become harder to forecast? *Journal of Money, Credit and banking*, 39(s1):3–33.

Stock, J. H. and Watson, M. W. (2012). Disentangling the channels of the 2007-2009 Recession. Technical report, National Bureau of Economic Research.

Svensson, L. E. (2006). Social value of public information: Comment: Morris and Shin (2002) is actually pro-transparency, not con. *The American Economic Review*, 96(1):448.

Tillmann, P. (2011). Strategic forecasting on the FOMC. *European Journal of Political Economy*, 27(3):547–553.

Uhlig, H. (1996). A law of large numbers for large economies. *Economic Theory*, 8(1):41–50.

Valev, N. T. and Carlson, J. (2003). Sources of dispersion in consumer inflation forecasts. *Applied Economics Letters*, 10(2):77–81.

Van Nieuwerburgh, S. and Veldkamp, L. (2009). Information immobility and the home bias puzzle. *The Journal of Finance*, 64(3):1187–1215.

Woodford, M. (2003). Imperfect common knowledge and the effects of monetary policy. *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, page 25.