Implications of Endogenous Cognitive Discounting*

James Moberly[†]

February 2022

Abstract

Cognitive discounting offers a plausible, tractable means of resolving numerous macroeconomic puzzles. The prevailing approach in the literature is to analyze and estimate models with exogenous cognitive discount factors. This paper uses a series of examples in a New Keynesian model to show that policy analysis and estimation results change considerably when discounting is endogenous. In terms of policy, endogenizing the discount factor significantly alters the determinacy condition, creates regime-dependence in the effects of changes in the Taylor rule parameters, and dramatically increases the benefits of average inflation targeting. In terms of estimation, endogenizing the discount factor resolves the weak identification found in models with exogenous discounting, leading to novel empirical results. In contrast to exogenous discounting models, my results suggest that indeterminacy cannot explain the Great Inflation. I also find that endogenous discounting offers an explanation for why the Phillips and IS curves appear flat during periods of macroeconomic stability.

JEL Classification: E30, E52, E70

*I would like to thank Guido Ascari and Martin Ellison for their invaluable advice and support throughout this project. I am grateful for helpful comments and suggestions from seminar participants at the University of Oxford. Financial support from the Oxford-Chellgren Graduate Scholarship is immensely appreciated.

[†]Department of Economics and University College, University of Oxford

1 Motivation

In a highly influential and important contribution, Gabaix (2020) demonstrates that introducing cognitive discounting into the new Keynesian framework resolves numerous macroeconomic puzzles. Gabaix supposes that agents pay limited attention to future deviations of macroeconomic variables from their steady states when making decisions today. This assumption introduces discount factors in front of the expectations terms in the IS and Phillips curves. This modification fixes a number of new Keynesian puzzles in a plausible, parsimonious, and tractable fashion; del Negro et al.'s (2015) forward guidance puzzle is resolved, the costs of the effective lower bound (ELB) are drastically reduced, and Ricardian equivalence is broken. The discount factors also relax the determinacy condition. As Cochrane (2016) notes, this could explain the lack of sunspot volatility in the ELB episode after the financial crisis.

Gabaix shows shows how one could in principle calculate an agent's optimal choice of cognitive discount factor, taking as given the model's dynamics, using the sparsity-based framework developed in two earlier papers.[1] He does not consider a model in which agents' choices of cognitive discount factor feed back into macroeconomic dynamics, creating a fixed point problem. Moreover, when analysing the model's properties, he takes the cognitive discount factor as given. As he puts it:

> The traditional New Keynesian model takes pricing frictions as given, and then studies their consequences. One can also endogenize the size of the pricing friction, but most of the analysis is most cleanly done by taking the pricing friction as given. Likewise, in this paper I take the degree of inattention as given, and study its consequences.

That makes sense in the context of Gabaix's paper, given that it already makes a far-reaching contribution to the literature. It has also been used in numerous papers that have used cognitive to study questions such as average inflation targeting, the causes of the Great Inflation, and medium-scale versions of the model.[2] Nonetheless, we know from other branches of the literature on bounded rationality in macroeconomics, such as Mackowiak and Wiederholt's (2015) business cycle model of rational inattention, that endogenous changes in how agents form expectations may be important for policy analysis. As such, it is important to assess whether the implications of Gabaix's model change qualitatively when attention is endogenized. As Cochrane (2016) argues, "because the paper is so important, its foundations matter".

In this paper, I take seriously the possibility that cognitive discounting is endogenous, and ask whether this changes policy analysis and the results obtained when estimating the model. I show through a series of examples using a new Keynesian model with endogenous discounting that the answer is a definitive "yes". I begin by extending Gabaix's derivations for calculating the optimal choice of cognitive discount factor to a richer set of models, that will allow me to assess endogenous cognitive discounting in models featuring indeterminacy, lagged terms, and regime switching. I then define an equilibrium choice of attention to the future, and analyze the existence of equilibrium in a simple version of the model.

I then show three theoretical results that emerge in the endogenous discounting models that are not present in models of exogenous discounting. First, I analyze the determinacy condition. Absent further assumptions, endogenizing the discount factor restores the original Taylor principle. That is, if the rational expectations Taylor principle is violated, then there always exists an equilibrium level of attention in which the economy is indeterminate. This may coexist alongside a determinate equilibrium. There are consequently two kinds of multiplicity; there may be multiple equilibrium levels of attention, and some of these allow multiple stable solutions to the model. Nonetheless, I show that the determinacy condition becomes path-dependent; if one starts in a low attention equilibrium, then it may not be possible to jump to the indeterminate equilibrium. Whether a determinate equilibriume exists depends strongly on the size of the shocks hitting the economy. This establishes a link between the size of the shocks hitting the economy and the implications of different monetary policy rules.

[1] See Gabaix (2014) and Gabaix (2017) for the details of the sparsity based framework.

[2] For example, see Ilabaca et al. (2020), Bundiato et al. (2021), Meggiorini and Milani (2021), and Meggiorini (2021).

Second, I show that endogenous discounting implies that the effects of changes in the Taylor rule parameters are regime dependent. Specifically, increases in the monetary policy response to inflation generate proportionally far larger effects on inflation volatility when the shocks hitting the economy are larger. Note that this result does not emerge under rational expectations or exogenous cognitive discounting. This finding results from two effects. First, the direct effect of changes in the policy response on inflation volatility is greater when expectations are closer to rationality. Second, a stronger policy response to inflation lowers firm attention, which in turn reduces inflation volatility. This finding may be of particular relevance to monetary policymakers handling the large increase in inflation that has followed the COVID-19 pandemic, and may help explain Andre et al.'s (2021) finding that households believe the current inflationary episode will be persistent.

Third, I analyze the benefits of average inflation targeting, which has been shown to mitigate the costs of the ELB under rational expectations. Previous analyses have noted that as average inflation targeting operates through affecting expectations of the future, its effects are muted under exogenous cognitive discounting (see Bundiato et al. (2021)). I show that the effects may be far larger under endogenous discounting. The reason is that average inflation targeting may push the economy to a far greater equilibrium level of discounting. This finding contributes to a broader literature assessing the effects of alternative monetary policy rules under bounded rationality.[3]

I then turn to empirical assessment of the model. I show that the exogenous discounting model suffers from weak identification throughout the parameter space; changes in the discount factors are difficult to distinguish from changes in the shock volatilities and Phillips and IS curve slopes. I show that endogenizing the discount factor resolves this problem by pinning down the discount factors in terms of the remaining model parameters.

I show two empirical applications of the endogenous discounting model. First, I estimate the model using US data for the Great Inflation and Great Moderation periods. Unlike previous analyses that used the exogenous discounting model (Ilabaca et al. (2020)), I find strong evidence against indeterminacy as a cause of the Great Inflation.[4] The endogenous discounting model also allows me to conduct counterfactuals that robust to Lucas' (1976) critque. Consistent with my theoretical analysis, I find that because the shocks hitting the economy during the Great Inflation were large, a stronger policy response to inflation would have substantially reduced inflation volatility. During the Great Moderation, by contrast, changes in the policy rule would have had proportionally far smaller effects. As well as demonstrating the importance of endogenising the discount factor when estimating the model, these findings contribute to the older "good luck" and "good policy" literatures regarding the Great Inflation.[5]

Second, I use simulations to show that endogenous discounting implies that an econometrician incorrectly assuming rational expectations would find that the Phillips and IS curves appear relatively flat during periods of macroeconomic stability. This occurs because agents discount more when shocks are small, implying that expectations provide less amplification. The effect of changes in interest rates on the output gap, and of changes in the output gap on interest rates, then decreases. An econometrician assuming rational expectations would incorrectly assess that the Phillips and IS curves had flattened. One can recover the correct slopes using an endogenous discounting model. An exogenous discounting model does not accurately recover the slope coefficients, because of weak identification. This may help explain the apparent flattening of the Phillips curve often observed in the data in recent decades.[6] This finding contributes to the literature suggesting that changes in expectation formation may affect identification

[3]See, for example, Bernanke et al.'s (2019) assessment of the "overshooting" problem using a model in which agents learn about macroeconomic dynamics using a vector autoregression.

[4]Another analysis of the Great Inflation and Great Moderation periods using exogenous discounting is Meggiorini (2021). In that paper, however, the consumer and firm discount factors are assumed to be equal. Although this assumption can help identification, one important conclusion from the empirical sections of this paper is that firm and consumer incentives to pay attention may differ very greatly.

[5]See, for example, Clarida et al. (2000) or Lubik and Schorfheide (2004) for the "good policy" view of the Great Inflation, and Fernandez-Villaverde et al. (2010) or Justiniano and Primiceri (2008) for the "good luck" hypothesis.

[6]See, for example, IMF (2013). A broader survey of this literature can be found in McLeay and Tenreyro (2019).

of the Phillips curve slope,[7] and offers a resolution to that problem.

Using endogenous discounting as an identification strategy relies on having estimates of the costs of paying attention. There are only a few studies investigating these costs empirically using microdata.[8] As such, the quantitative empirical findings are necessarily tentative. However, I do provide a means of externally validating the findings. I calculate degree of information rigidity in expectation formation that the estimated model implies for different periods, and compare that to the level of information rigidity found in SPF data by Coibion and Gorodnichenko (2015). I find that my estimates for the Great Moderation period match these estimates from expectations data fairly closely.

The two most closely related papers are Chau (2020) and Lubik and Marzo (2021). Chau uses Gabaix's sparsity based framework to endogenize attention to contemporaneous macroeconomic variables. The model generates time variation in the slope of the Phillips curve, which he applies to the "missing disinflation" puzzle. This form of inattention differs from the cognitive discount factor assumed by Gabaix; Chau assumes that agents discount macroeconomic variables at all horizons equally. Cognitive discounting, conversely, is dynamic; realizations of the endogenous variables in the distant future are more heavily discounted. It is this feature that generates determinacy in Gabaix's model even under a passive interest rate policy, as well as resolving the forward guidance puzzle, the costs of the effective lower bound, and so on. Chau does not assume dynamic discounting, stating that it is unnecessary for explaining the time variation in the slope of the Phillips curve.[9] Lubik and Marzo (2021) investigate fiscal policy in a model with cognitive discounting. They do include a brief section in which they calculate the equilibrium choice of attention for a particular calibration. However, they only calculate this optimal choice of attention for one calibration of the model, whereas this paper is about how the equilibrium level of attention changes with changes in policy rules, shock processes, and so on. They also do not look at different monetary policy rules under endogenous attention (which comprise the three theoretical applications in this paper), or identification and estimation under endogenous attention. Moreover, the technical contributions developed in the earliest version of this paper (see Moberly (2020)) are (i) to extend Gabaix's method of calculating optimal attention to allow the calculation of attention equilibrium in a richer class of models, rather than using Gabaix's analytical derivations for a simple model, and (ii) to prove properties about the existence of determinate and indeterminate equilibria.

My paper also relates to a broader literature that offers several differnt means of microfounding discounting in new Keynesian models. Examples include McKay et al.'s (2016) model with uninsurable income risk, and Bilbiie's (2021) model which features discounting when inequality is procyclical. My results show that how the discounting is microfounded matters; as such, adjudicating between different sources of discounting remains a critical objective for future research.

I also contribute to the wider literature incorporating bounded rationality into DSGE models. These approaches include learning (see Evans and Honkapohja (2001)), heuristic switching (see Hommes (2018)), and the level-k thinking (see Farhi and Werning (2019)). Other deviations from rationality are surveyed by Woodford (2013). Perhaps most closely related are the sticky information approach studied by Mankiw and Reis (2002, 2006), and microfounded by Reis (2006a, 2006b), and the rational inattention approach proposed by Sims (2003) and developed in a DSGE setting by Mackowiak and Wiederholt (2009, 2015). These two approaches share a similarity in that agents trade off a cost of being more rational against the losses associated with deviating from rationality. Nonetheless, the implications of rational inattention and cognitive discounting differ, for example in terms of their impact on the determinacy condition, so my paper is differentiated from these studies. My paper does, however, reinforce the conclusion that modelling deviations from rationality as endogenous phenomena is important for policy analysis.

[7]See, for example, Lansing and Jorgensen's (2022) study, which uses a signal extraction model. The idea that more anchored inflation expectations could explain the smaller observed pass-through from output to inflation is also discussed in IMF (2013), and also in Ball and Mazumder (2015).

[8]One example is Ganong and Noel (2017).

[9]See footnote 5 in Chau (2020)

2 Simple Model and Determinacy Condition

This section presents a simple version of the endogenous attention model, defines the concept of attention equilibrium, and investigates the existence of indeterminate and determinate equilibria. To simplify the presentation, in this section I use a greatly simplified model which abstracts from firms' expectations, from fundamental shocks, and from intrinsic sources of persistence. I will introduce a richer model in section 3.

I begin by assuming that inflation is just a linear function of output. That is, there is no expectation term in the Phillips curve, nor are there cost-push or mark-up shocks.

$$\pi_t = \kappa y_t \tag{1}$$

I assume a very simple monetary policy rule, whereby the nominal rate responds only to inflation. Again, there is no fundamental shock.

$$i_t = \phi_\pi \pi_t \tag{2}$$

For now, I assume no government. This is a closed economy model with no capital investment, and so in equilibrium consumption equals output.

$$c_t = y_t \tag{3}$$

As in Gabaix (2020), I combine the consumer's intertemporal optimization condition under rationality with the budget constraint, and then add a discount factor $m_c \in [0,1]$.

$$c_t = \sum_{h\geq 0} (\beta m_c)^h ((1-\beta)E_t y_{t+h} - \beta\sigma(E_t i_{t+h} - E_t \pi_{t+h+1})) \tag{4}$$

This formulation stems from the h period ahead forecast under bounded rationality, denoted $\tilde{E}_t$ being a discounted version of the rational expectation E_t. For inflation, for example:

$$\tilde{E}_t \pi_{t+h} = (m_c)^h E_t \pi_{t+h} \tag{5}$$

Using the goods market clearing condition, and re-arranging, one obtains the aggregate Euler equation:

$$y_t = M_c E_t y_{t+1} - \sigma(i_t - E_t \pi_{t+1}) \tag{6}$$

Here, uppercase M_c denotes the aggregate cognitive discount factor. Although $M_c = m_c$ because there is a representative agent, the distinction is important when attention is endogenous. Agents will take the aggregate level of discounting M_c as given when choosing their individual cognitive discount factor m_c.

2.1 Exogenous attention

Taking as given the level of attention, the model can be re-written as:

$$y_t = \delta E_t y_{t+1} \tag{7}$$
$$\delta = \frac{M_c + \kappa\sigma}{1 + \kappa\sigma\phi_\pi}$$

If $\delta < 1$ then the model has a unique stable solution: $y_t = 0$ for all t. This requires:

$$\phi_\pi > 1 - \frac{1 - M_c}{\kappa\sigma} \tag{8}$$

If $M_c = 1$, so that agents pay full attention, then expectations are rational, and so one obtains the familiar condition that $\phi_\pi > 1$ ensures determinacy. As M_c decreases, so agents discount the future more, the determinacy condition relaxes. As Gabaix notes, if κ and σ are sufficiently small, then even moderate discounting ensures determinacy for any $\phi_\pi \geq 0$.

If $\delta > 1$, then the model is indeterminate. Specifically, y_t follows an autoregressive process:

$$y_t = \delta^{-1} y_{t-1} + \eta_t \tag{9}$$
$$\eta_t = y_t - E_{t-1} y_t$$

Absent fundamental shocks, the expectation error is just given by a sunspot shock, which I denote ζ_t. The sunspot shock must be mean zero, but the model places no restrictions on higher moments of ζ_t. I assume that it has constant variance σ_ζ^2.[10]

If $\eta_t = \zeta_t$, then the stationary distribution of y_t then has variance:

$$\sigma_y^2 = \frac{\sigma_\zeta^2}{1 - \delta^{-2}} \tag{10}$$

Notice that at the very edge of the indeterminacy region, as δ approaches unity from above, the stationary variance of output rises becomes unboundedly large. Why? In the indeterminacy region, sunspot shocks are mean-reverting, which is why we do not rule them out when finding stable solutions to the model. In the determinacy region they are explosive, which is why they are ruled out. At the boundary, they have a unit root. As one approaches the boundary, then, the sunspot shocks approach a unit root and so generate unboundedly large variance in output over a sufficiently long time horizon.

2.2 Endogenizing attention

Let's now endogenize the level of attention. As in Gabaix's microfoundation, suppose that the consumer trades off a loss from inattention and a cost of paying attention. I suppose that the consumer takes the aggregate cognitive discount factor M_c as given. The consumer's problem is:

$$\min_{m_c \in [m_{c,d}, 1]} L_c(m_c, M_c, \boldsymbol{\chi}) + C(m_c, \boldsymbol{\xi_c}) \tag{11}$$

Here, $m_{c,d}$ is some "default", minimum level of attention. It can be set to zero. The loss from inattention, compared to paying full attention $m_c = 1$, is $L_c(m_c, M_c, \boldsymbol{\chi})$. The utility loss is necessarily zero if the consumer pays full attention, hence $L_c(1, M_c, \boldsymbol{\chi}) = 0$. The loss depends on (i) other consumer's choices of attention M_c, and (ii) the remaining model parameters $\boldsymbol{\chi}$. The cost of paying attention $C(m_c, \boldsymbol{\xi_c})$ will typically be assumed to be strongly monotonic in m_c, and it depends on a vector of parameters $\boldsymbol{\xi_c}$ which determine the shape and scale of the attention cost. I denote the optimal choice of m_c as:

$$g_c(M_c, \boldsymbol{\chi}, \boldsymbol{\xi_c}) = \arg \min_{m_c \in [m_{c,d}, 1]} L_c(m_c, M_c, \boldsymbol{\chi}) + C(m_c, \boldsymbol{\xi_c}) \tag{12}$$

I then define an *attention equilibrium* as a fixed point of this mapping. That is, a choice of attention that implies a set of dynamics which in turn justify that choice of attention as optimal.

Definition: An *equilibrium choice of attention* is a choice of attention $M_c(\boldsymbol{\chi}, \boldsymbol{\xi_c})$ such that:

$$M_c(\boldsymbol{\chi}, \boldsymbol{\xi_c}) = g_c(M_c(\boldsymbol{\chi}, \boldsymbol{\xi_c}), \boldsymbol{\chi}, \boldsymbol{\xi_c}) \tag{13}$$

Notice that this setup goes slightly further than the proposed endogenization in section VIII.B of Gabaix (2020). Gabaix imposes specific functional forms for how agents approximate the loss from not paying attention and on the cost of paying attention. He then solves for the optimal choice of attention (in the case with just technology shocks, no intrinisic persistence, and determinacy), taking as given the model dynamics. Here, I note that the equilibrium dynamics depend on the choice of attention, set up a fixed point mapping, and note that in equilibrium the choice of attention must be a fixed point of this mapping. This is perhaps closer to the idea used in Gabaix (2014) to solve for the equilibrium price in a

[10] The assumption that the variance does not change over time is innocuous. Less innocuous is the assumption that it does not depend on δ. That is, there is no tendency for sunspot shocks of different persistence to have different variance. In particular, for the derivation below, one would have to assume that σ_ζ^2 does not tend to zero as δ approaches unity from above.

microeconomic general equilibrium setting, where the equilibrium price is found using a fixed point.

For most of this section, I will assume a particular form for $L_c(m_c, M_c, \boldsymbol{\chi})$ and $C(m_c, \boldsymbol{\xi_c})$, as Gabaix suggests. However, the results apply much more generally, and below I will give sufficient conditions on $L_c(m_c, M_c, \boldsymbol{\chi})$ and $C(m_c, \boldsymbol{\xi_c})$ for them to hold, and investigate how they work in other specific assumptions. Specifically, I follow Gabaix in assuming that agents approximate losses using a second-order approxiation of the value function V:

$$L_c(m_c, M_c, \boldsymbol{\chi}) = \frac{1}{2}(1-m_c)^2 V_{cc} E\left[\left(\frac{\partial c}{\partial m_c}\right)^2\right] \tag{14}$$

I follow Gabaix in allowing agents to approximate the derivative of consumption with respect to m_c by the derivative at the "default" level of attention $m_{c,d}$. I assume that the cost of increasing m_c above the default is linear:

$$C(m, \boldsymbol{\xi_c}) = \mathcal{K}|m_c - m_{c,d}| \tag{15}$$

Throughout, I use the scale-free attention cost discussed in Gabaix (2017), which is invariant to transformations of the utility function. The cost of cognition $\mathcal{K}$ is:[11]

$$\mathcal{K} = k_c^2 (c(m_d))^2 |V_{cc}| \tag{16}$$

Here, k_c is the scale-free cost of attention. The intuition for k_c is that if $k = 1.0$, for example, then agents pay attention to variables that on average make more than a 1% difference to their optimal action at the typical scale. $c(m_d)$ is the choice of consumption at the default level of attention. As this never differs by more than a few percentage points from the steady-state level of consumption, which I scale to unity, I approximate by dropping this term, which simplifies the expressions. Under these assumptions, the choice of attention is:

$$g_c(M_c, \boldsymbol{\chi}, \boldsymbol{\xi_c}) = \max\left(1 - \frac{k^2}{E\left[\left(\frac{\partial c}{\partial m}\right)^2\right]}, m_{c,d}\right) \tag{17}$$

To find $E\left[\left(\frac{\partial c}{\partial m}\right)^2\right]$, one can simply differentiate equation (4) with respect to m_c.

I will denote by $\tilde{m}_c(\boldsymbol{\chi})$ the level of attention that gives $\delta = 1$, taking as given the remaining model parameters. That is:

$$\tilde{m}_c(\boldsymbol{\chi}) = 1 + \kappa\sigma(\phi_\pi - 1) \tag{18}$$

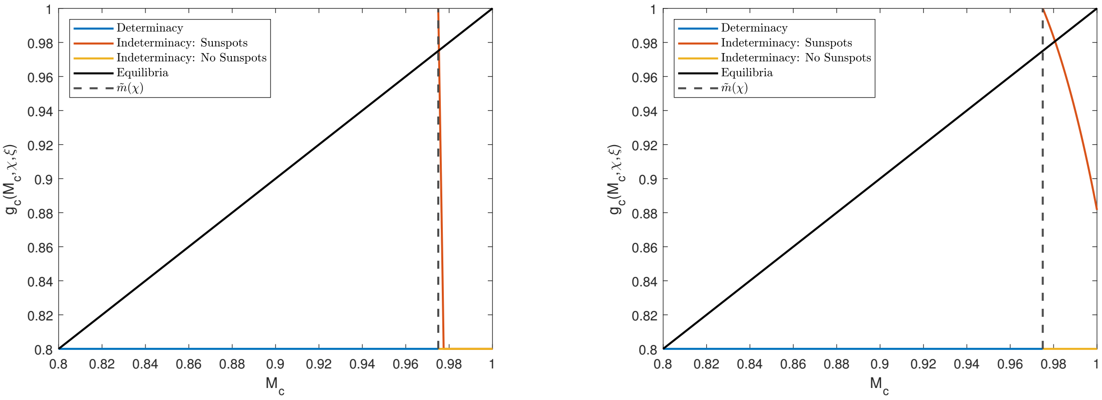
In the example below, I will assume that $\phi_\pi < 1$, so that $\tilde{m}_c(\boldsymbol{\chi}) < 1$. That is, I assume that the rational expectations Taylor principle is violated, and the economy is indeterminate if agents pay full attention. I will also assume that $\tilde{m}_c(\boldsymbol{\chi}) > m_{c,d}$, so the economy is determinate at the default, minimum level of attention.

If $M_c < \tilde{m}(\boldsymbol{\chi})$, so that $\delta < 1$, then there is no incentive to pay attention: $c_t = 0$ in all t irrespective of the choice of m, so agents always choose $M_c = m_{c,d}$. Provided that $\delta < 1$ when $M_c = m_{c,d}$, then this is an equilibrium.

Suppose instead that $\delta > 1$. What are the incentives to pay attention? We have:

$$c_t = \sum_{h \geq 0} (\beta m_c)^h ((1-\beta) E_t y_{t+h} - \beta\sigma E_t i_{t+h} + \beta\sigma E_t \pi_{t+h+1}) \tag{19}$$

[11] See Gabaix (2017), equation (75).



(a) Small sunspot shocks

(b) Large sunspot shocks

Figure 2.1: Attention equilibria

After some algebra, one finds that the expectation of the square of the derivative (evaluated at the default) is:

$$E[(c_{m,t})^2] = \frac{\beta^2\delta^{-2}}{(1-\beta m_{c,d}\delta^{-1})^4}(1-\beta-\beta\sigma\kappa(\phi_\pi-\delta^{-1}))^2\sigma_y^2 \quad (20)$$

$$= \frac{\beta^2\delta^{-2}}{(1-\beta m_{c,d}\delta^{-1})^4}(1-\beta-\beta\sigma\kappa(\phi_\pi-\delta^{-1}))^2\frac{\sigma_\zeta^2}{1-\delta^{-2}} \quad (21)$$

Figure 2.1 plots the resulting g_c mapping in two calibrated examples. Throughout this section, the calibration used has: $\beta = 0.99$, $\kappa = 0.10$, and $\sigma = 0.50$. I assume that $\phi_\pi = 0.50$, so that the rational expectations Taylor Principle is violated and $\tilde{m}(\chi) < 1$. I assume that $m_{c,d} = 0.80$, and $k_c = 1.0$. Attention equilibria are on the 45 degree line, where a given level of aggregate attention maps into the same choice of individual (and hence aggregate) attention. In the determinate region, the g_c mapping is shown by the blue line. Absent fundamental shocks, output volatility and hence incentives to pay attention are zero. So, the blue line remains at the default level of attention, whatever is the aggregate choice of attention. That gives an attention equilibrium at the default level of attention of 0.80. In the indeterminate region, I show the g_c mapping using the red line. For comparison, the yellow line shows the case where the sunspot shock variance is equal to zero. Here, the no sunspot shock case implies zero output volatility, no incentives to pay attention, and so an optimal choice of attention at the default. The red line, on the other hand, shows the case with sunspot shocks. In panel (a), I set $\sigma_\zeta = 0.2$, while in panel (b), I use $\sigma_\zeta = 1.0$. Notice that as $m_c \to \tilde{m}_c(\boldsymbol{\chi})$, δ approaches unity from above, and the variance of output rises without limit. Incentives to pay attention become unboundedly large, and so the optimal choice of attention (shown by the red line) approaches unity. The continuity of $g_c(M_c, \boldsymbol{\chi}, \boldsymbol{\xi_c})$ in M_c within the indeterminacy region then ensures that an indeterminate equilibrium always exists. This situation holds whenever the rational expectations Taylor principle fails. Our specific example always allows for a determinate equilibrium to exist. As such, the model features two kinds of multiplicity. First, there are multiple equilibrium choices of attention. Second, one of those equilibrium choices of attention allows for multiple stable solutions.

2.3 Do the functional forms matter?

What assumptions are needed on L and C to ensure that an indeterminate equilibrium always exists whenever $\tilde{m}(\boldsymbol{\chi}) \in [m_{c,d}, 1]$? Suppose that $\tilde{m}_c(\boldsymbol{\chi}) \in [m_{c,d}, 1]$. Then a sufficient set of conditions for an attention equilibrium to exist in the interval $(\tilde{m}_c(\boldsymbol{\chi}), 1]$ is:

- $L_c(m_c, M_c, \boldsymbol{\chi})$ and $C(m_c, \boldsymbol{\xi})$ are continuous in M_c between $\tilde{m}_c(\boldsymbol{\chi})$ and unity.
- $L_c(m_c, M_c, \boldsymbol{\chi}) = 0$ when $m_c = 1$, and $L_c(m_c, M_c, \boldsymbol{\chi}) \to \infty$ when $Var(E_t\pi_{t+1}) \to \infty$ if $m_c < 1$.
- $C(m_c, \boldsymbol{\xi_c})$ is finite for all $m_c \in [m_{c,d}, 1]$.

The first condition is technical. The second assumption simply requires that losses from inattention are (i) zero under rationality, and (ii) become unbounded if the agent does not pay attention to expected future fluctuations in output and those fluctuations are unboundedly large. The third assumption requires a finite cost of the agent being rational. Under these assumptions, $g_c(M_c, \boldsymbol{\chi}, \boldsymbol{\xi_c})$ approaches unity as M_c approaches $\tilde{m}_c(\boldsymbol{\chi})$ from above, because losses approach infinity for any choice of m_c less than unity. Moreover, the assumptions imply that $g_c(M_c, \boldsymbol{\chi}, \boldsymbol{\xi_c})$ is continuous in M_c by Berge's maximum theorem. That gives the result that an indeterminate equilibrium must exist if $\tilde{m}_c(\boldsymbol{\chi}) \in [m_{c,d}, 1]$. The result that an indeterminate equilibrium always exists is consequently not driven by the specific functional forms used for L and C, but applies much more generally.

2.4 An equilibrium refinement

The indeterminacy result above relied on a quite specific assumption; I showed that there always exists an indeterminate equilibrium, provided that the economy has been in the indeterminate state for an arbitrarily long time, so that the variance of output was given by:

$$\sigma_y^2 = \frac{\sigma_\zeta^2}{1-\delta^{-2}}$$

From now on, I'll refer to this equilibrium as a *long-run* equilibrium, to denote that it is an equilibrium that exists if the economy has been in the indeterminate region already for a long time.

Suppose instead that the economy was in the determinate state in time $t-1$. Then $y_{t-1} = 0$. I now ask: could the economy make an unexpected jump into an indeterminate equilibrium in time t? In time t, output is given by:

$$y_t = \delta^{-1} y_{t-1} + \zeta_t \tag{22}$$

The variance of output is:

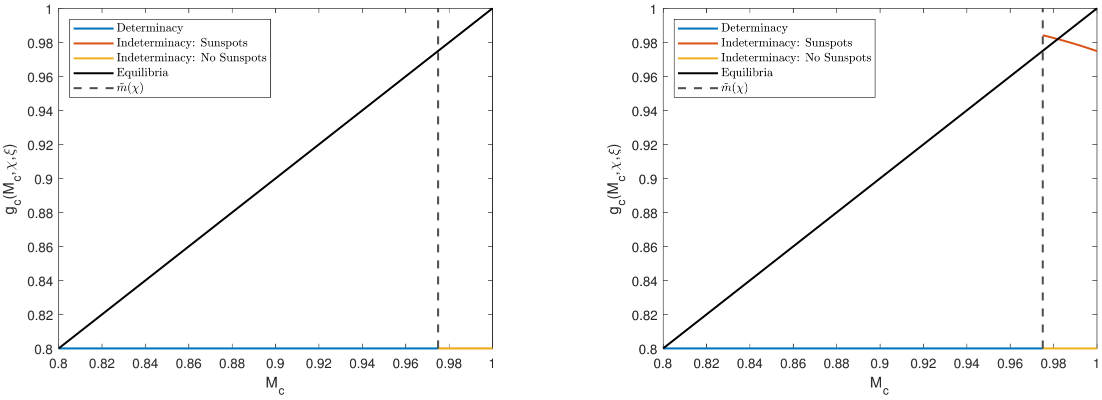
$$\begin{aligned} Var(y_t) &= \delta^{-2} Var(y_{t-1}) + \sigma_\zeta^2 && (23)\\ &= \sigma_\zeta^2 && (24) \end{aligned}$$

Then the critical term in calculating the g_c mapping is:[12]

$$\begin{aligned} E[(c_{m,t})^2] &= \frac{\beta^2\delta^{-2}}{(1-\beta m_{c,d}\delta^{-1})^4}(1-\beta-\beta\sigma\kappa(\phi_\pi - \delta^{-1}))^2\sigma_y^2 && (25)\\ &= \frac{\beta^2\delta^{-2}}{(1-\beta m_{c,d}\delta^{-1})^4}(1-\beta-\beta\sigma\kappa(\phi_\pi - \delta^{-1}))^2\sigma_\zeta^2 && (26) \end{aligned}$$

Note that this term does *not* become unboundedly large as δ approaches unity from above. Hence although there is a discontinuity in the g_c mapping at $\tilde{m}$, it is a small discontinuity, rather than taking the g mapping up to unity. Indeed, if the variance of the sunspot shocks is small enough, then attention may remain at its default level as one crosses the edge of the determinacy region. This situation is shown in figure 2.2(a), which uses the same calibration as figure 2.1(b). Notice that under either level of attention cost, the output of the g_c mapping remains at its default level at the boundary. As such, there is no indeterminate equilibrium. So, if the economy starts in the determinate equilibrium, it will only be able to reach the indeterminate equilibrium if the variance of the sunspot shocks is sufficiently large. Consequently, the calibration considered in 2.1(b) and 2.2(a) allows an indeterminate equilibrium to be sustained if the model has already been in that equilibrium for an arbitrarily long time, but does not allow the economy ever to switch into the indeterminate state. In this sense, the long-run indeterminate equilibrium shown in figure 2.1(b) is *unattainable*.

[12] Strictly, this simplifies a little by supposing that agents assume that if a particular M_c prevails in time t, then it will always prevail. Otherwise, the δ would be expected to change over time, which would make the calculation more complex. This simplification makes the derivations that follow more tractable.



(a) Large sunspot shocks

(b) Very, very large sunspot shocks

Figure 2.2: Attention equilibria, assuming determinacy in the previous period

The determinacy condition of the model is consequently dynamic, and path dependent. Starting off in the determinate regime means that the model is likely to remain there. Looking at long-run equilibria restored a version of the rational expectations Taylor principle; whenever the rational expectations Taylor principle fails, a long-run indeterminate equilibrium exists. If one looks only at the equilibria that can be reached, however, then policy rules which break the rational expectations Taylor principle will often only allow for a determinate equilibrium to exist.

Of course, the model does not place restrictions on the size of the sunspot shocks. One can always find a sufficiently large σ_ζ such that an indeterminate equilibrium exists. This situation is shown in figure 2.2(b), in which I set the sunspot shock variance to be ten times larger, so that $\sigma_\zeta = 10.0$. This generates enormous volatility in consumption; the expected absolute value of the derivative of consumption with respect to attention is approximately 7% in the period after the switch. That is sufficient to induce agents to attention even immediately after a switch, which allows an indeterminate equilibrium will exist in the period after the switch; notice the intersection of the red line in figure 2.2(b) with the 45 degree line. Nonetheless, one might be skeptical about the possibility of such large sunspot shocks; these would require an extraordinary degree of coordination between agents.

What happens if an indeterminate equilibrium does exist in the period after the switch? I now introduce some additional notation. First, denote $M_{c,t+\tau}$ as the aggregate level of attention chosen in period $t+\tau$, and $\delta_{t+\tau}$ as δ implied by this level of attention. Now in period $t+\tau$, supposing that the economy remains within the indeterminate region:

$$Var(y_{t+\tau}) = \delta_{t+\tau-1}^{-2} Var(y_{t+\tau-1}) + \sigma_\zeta^2 \tag{27}$$

It must be that the variance of $y_{t+\tau}$ is at least as great as the variance of y_t, which means that the resulting g_c will be at least as great for any M_c. This means that if an indeterminate equilibrium exists in the first period after the switch, it will exist in all subsequent periods. In the simple example considered so far, then, if an indeterminate equilibrium exists in the first period after the switch, then it will exist in all subsequent periods, with attention drifting up towards the long-run equilibrium over time. To formally define the notion of equilibrium attainability, I first define a new mapping $g_c^{att}(M_c, \boldsymbol{\chi}, \boldsymbol{\xi_c})$, which is identical to the original g_c mapping given in (17) but using the derivative of consumption with respect to attention in (26). I then define an indeterminate equilibrium given by the long-run g_c mapping as attainable only if there also exists an indeterminate equilibrium of the short-run g_c^{att} mapping.

Definition: An equilibrium choice of attention $M_c(\boldsymbol{\chi}, \boldsymbol{\xi_c})$that lies in the indeterminate region is *attainable* if an only if there exists some $M_c^* \in [\tilde{m}(\boldsymbol{\chi}, 1]$ such that:

$$M_c^* = g_c^{att}(M_c^*, \boldsymbol{\chi}, \boldsymbol{\xi_c}) \tag{28}$$

2.5 Fundamental shocks

So far I abstracted entirely from fundamental shocks. That meant that in the determinate state, output and inflation were always zero, and hence there were no incentives to pay attention, guaranteeing the existence of a determinate attention equilibrium at the default level of attention whenever $m_{c,d} < \tilde{m}(\boldsymbol{\chi})$.

Let's now introduce a fundamental shock. Suppose that there is a policy shock, so that the nominal interest rate is:

$$i_t = \phi_\pi \pi_t + v_t \tag{29}$$

$$v_t = \rho v_{t-1} + \varepsilon_t^v \tag{30}$$

$$\varepsilon_t^v \sim N(0, \sigma_{\varepsilon,v}^2)$$

The model now reduces to:

$$y_t = \delta E_t y_{t+1} - \tilde{v}_t \tag{31}$$

$$\tilde{v}_t = \frac{\sigma}{1 + \kappa\sigma\phi_\pi} v_t \tag{32}$$

I denote the variance of $\tilde{v}_t$ as $\tilde{\sigma}_v^2$.

Under determinacy, the model solution is:

$$y_t = -\frac{1}{1 - \delta\rho} \tilde{v}_t \tag{33}$$

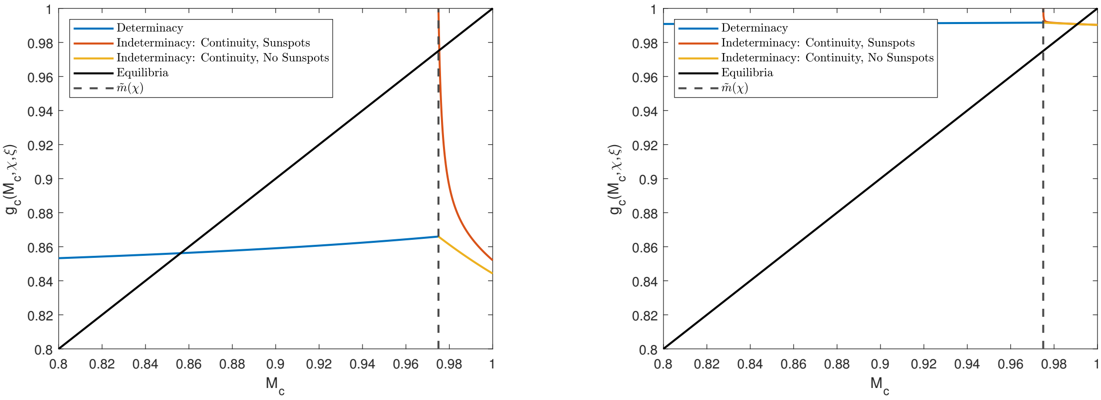
The consumption choice is linear in $\tilde{v}_t$, because all of the endogenous variables and their expectations are just linear functions of $\tilde{v}_t$. Hence the derivative of consumption with respect to m_c is also linear in $\tilde{v}_t$. The expectation of the square of this quantity is then just linear in $\tilde{\sigma}_v^2$ and hence in $\sigma_{\varepsilon,v}^2$.

Now suppose that $\tilde{m}(\boldsymbol{\chi}) < 1$, so that the economy is indeterminate under rationality. For any aggregate M_c within the determinate region, there exists some sufficiently large $\sigma_{\varepsilon,v}^2$ such that the implied individual choice of m_c approaches unity. That will ensure there is no determinate equilibrium. As such, the existence of determinate equilibrium under any policy rule with $\phi_\pi < 1$ depends critically on the size of the shocks hitting the economy. This situation is shown in figure 2.3(a) and (b). These figures assume that $\rho = 0.8$. 2.3(a) assumes that $\sigma_{\varepsilon,v} = 0.5$, while 2.3(b) assumes that $\sigma_{\varepsilon,v} = 2.0$. Notice the difference in the blue line, that shows the optimal choice of attention in the determinate region. In 2.3(a), with the smaller shocks, incentives to pay attention are sufficient to move the g_c mapping above the default level. However, a determinate equilibrium still exists at the intersection of the blue line with the 45 degree line. In 2.3(b), however, the fundamental shocks are larger. The g_c mapping in the determinate region is shifted upwards, to the extent that no equilibrium exists in the determinate region.

What does the long-run equilibrium look like in the indeterminate region? To pin this down, one has to not only specify the sunspot shock variance, but also the response of the endogenous variables to the fundamental shocks. The latter is not pinned down in the indeterminacy region, but one has to specify a solution in order to calculate the attention level. I will throughout use Lubik and Schorfheide's (2003, 2004) continuity solution. In this simple example, that means that at the very edge of the indeterminacy region, expectation errors are given by:

$$\eta_t = \frac{1}{1 - \rho} \varepsilon_t^v + \zeta_t \tag{34}$$

The reason is that this is the response at the boundary under determinacy (plus a sunspot shock), so the responses to fundamental shocks are continuous as one crosses the boundary. This assumption gives a finite variance for output and the boundary of the indeterminacy region if and only if $\sigma_\zeta^2 = 0$. This case is shown by the yellow lines in figures 2.3(a) and (b). Because the responses to the fundamental shocks are continuous as one crosses the boundary of the determinacy region, and there are no sunspot



(a) Small fundamental shocks, small sunspot shocks (b) Large fundamental shocks, small sunspot shocks

Figure 2.3: Attention equilibria with fundamental shocks

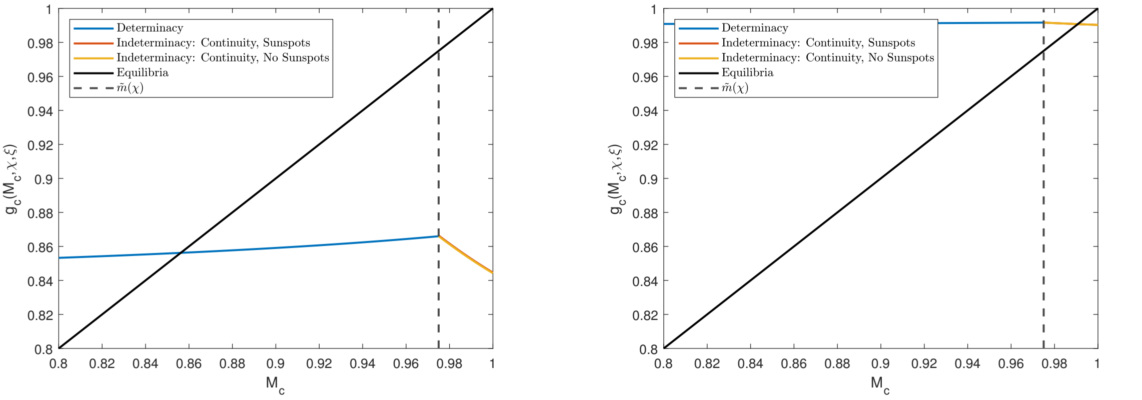
shocks, then the g_c mapping is continuous at the boundary, and there is no jump from the blue to the yellow line. However, if one deviates even slightly from the continuity solution, or if the sunspot shock variance is non-zero, then one must have an infinite output variance at the boundary, meaning that a long-run indeterminate equilibrium exists. The red line shows the g_c mapping when the sunspot shock variance is non-zero. Once again, the cumulation of sunspot shocks over time ensures that the g_c mapping approaches unity as the boundary of the indeterminate region. This effect in turn ensures that a long-run indeterminate equilibrium is sure to exist because of the continuity of the g_c mapping in M_c.

The attainability refinement is a more complex here. The reason is that when the economy jumps into the indeterminate region, one has to specify (i) the point at which it jumps to, in terms of the fundamental shocks plus any sunspot shock that hits in the period, and (ii) the subsequent effect of ε_t^v on the expectation errors η_t. Although there are many possible ways in which the economy can respond to the fundamental shocks, one has to specify a solution in order to calculate the appropriate attention level. I will again use Lubik and Schorfheide's (2003) "continuity solution". Right at the boundary, the continuity solution implies that:

$$y_t = \frac{1}{1-\rho}\tilde{v}_t + \zeta_t \tag{35}$$

The reader might then wonder whether that implies that the expectation error in time t is predictable at time $t-1$. Recall, though, that I am only entertaining the possibility of unexpected switches from one equilbrium to the other. One could also suppose that agents expect some probability of a switch between the two equilibria, but that is beyond the scope of this paper. Notice that at this solution for output, output in time t has a finite variance, and thus the g_c mapping at the boundary will be less than unity. As such, there is no guarantee that an indeterminate equilibrium will exist, although it may if the sunspot shock variances or the fundamental shock variances are sufficiently large. The other point to note in terms of existence of equilibrium is that there is always going to be some equilibrium, be it determinate or indeterminate. Under the continuity solution plus a sunspot shock, the variance of output and hence incentives to pay attention are always at least as great on the indeterminate side of the boundary. That upwards jump ensures that some equilibrium will always exist.

I show the same calibrations as in figure 2.3 in figure 2.4, but this time assuming that the economy was in the determinate region in the previous period. Notice that an indeterminate equilibrium does not exist in 2.4(a), where the fundamental shock variances are small, but does exist in 2.4(b), when the fundamental shock variances are large. In 2.4(a), there is a unique equilibrium level of attention, which admits only one stable solution. The model does not then suffer from multiplicity. In 2.4(b), there is still a unique equilibrium level of attention, but that level of attention admits many stable solutions. As such, under the attainability refinement, the determinacy of the model depends strongly on the fundamental shock variances. Larger fundamental shocks increase incentives to pay attention and so push the model



(a) Small fundamental shocks, small sunspot shocks (b) Large fundamental shocks, small sunspot shocks

Figure 2.4: Attention equilibria with fundamental shocks, assuming determinacy in previous period

towards indeterminacy.

Of course, one could always find some point that the economy could jump to in time t such that an indeterminate equilibrium exists. For example, one could just specify that the economy jumps to the point $y_t = A\tilde{v}_t + \zeta_t$ with an arbitrarily large A, which would give an arbitrarily large $Var(y_t)$. However, it seems *a priori* unlikely that agents would coordinate on such a solution, given how large an expectation error it requires compared to $E_{t-1}y_t$. That is why using the continuity solution seems (to me) to be the most plausible solution in this scenario.

3 A Richer Model

For simplicity, I derived the theoretical results in a very simplified environment. I now endogenize attention in a richer model. The Phillips curve, now includes firm expectations, and allows firms to be inattentive. I also incorporate fundamental shocks and interest rate smoothing.

My objective is to endogenize both firm and consumer attention, to obtain:

$$\pi_t = \beta M_f(\boldsymbol{\chi}, \boldsymbol{\xi}) E_t \pi_{t+1} + \kappa x_t + \eta_t \tag{36}$$

$$x_t = M_c(\boldsymbol{\chi}, \boldsymbol{\xi}) E_t x_{t+1} - \sigma(i_t - E_t \pi_{t+1}) + \tilde{z}_t \tag{37}$$

$$i_t = \rho_i i_{t-1} + (1 - \rho_i)(\phi_\pi \pi_t + \phi_x x_t) + \varepsilon_t^v \tag{38}$$

$$\eta_t = \rho_\eta \eta_{t-1} + \varepsilon_t^\eta \tag{39}$$

$$z_t = \rho_z z_{t-1} + \varepsilon_t^z \tag{40}$$

ε_t^η, ε_t^z, and ε_t^v are assumed to be Gaussian, with mean zero and variances $\sigma_{\varepsilon\eta}^2$, $\sigma_{\varepsilon z}^2$, and $\sigma_{\varepsilon v}^2$ respectively. Note that here I use x_t to denote the output gap, whereas $\hat{y}_t$ denotes the deviation of output from its steady state. However, as I motivate demand shocks as discount factor rather than technology shocks, the two are equal.

The determinacy condition is given by:

$$\phi_\pi + \frac{(1 - \beta M_f(\boldsymbol{\chi}, \boldsymbol{\xi}))}{\kappa}\phi_x + \frac{(1 - \beta M_f(\boldsymbol{\chi}, \boldsymbol{\xi}))(1 - M_c(\boldsymbol{\chi}, \boldsymbol{\xi}))}{\kappa\sigma} > 1 \tag{41}$$

This is the same as the condition derived by Gabaix (2020) for the model without interest rate smoothing.

I assume constant returns to scale, and Calvo pricing, so that κ is given by:

$$\kappa = (\gamma + \phi)\frac{(1 - \beta\theta)(1 - \theta)}{\theta} \tag{42}$$

The slope of the IS curve σ is given by γ^{-1}, where γ denotes the coefficient of relative risk aversion. θ is the Calvo pricing parameter, and ϕ is the inverse of the Frisch elasticity. Following Gabaix, the aggregate cognitive discount factors in this model M_f and M_c are related to the individual choices of cognitive discount factors m_f and m_c as follows:

$$M_f(\boldsymbol{\chi}, \boldsymbol{\xi}) = m_f(\boldsymbol{\chi}, \boldsymbol{\xi}) \left(\theta + (1-\theta) \frac{1-\beta\theta}{1-\beta m_f(\boldsymbol{\chi}, \boldsymbol{\xi})\theta} \right) \tag{43}$$

$$M_c(\boldsymbol{\chi}, \boldsymbol{\xi}) = m_c(\boldsymbol{\chi}, \boldsymbol{\xi}) \tag{44}$$

The more complex relationship between M_f and m_f arises because of Calvo pricing. The reader is referred to Gabaix (2020) for the derivation. Note that $\boldsymbol{\xi}$ comprises the attention function parameters for both firms and consumers. Because there may be strategic complementarities or substitutabilities between agents' choices of attention, firm attention may be affected by the consumer attention cost, and vice-versa. Formally, the g mapping now maps a vector $\boldsymbol{M}$, which comprises M_f and M_c, into a set of macroeconomic dynamics, through to choices of individual attention m_f and m_c, and finally into the implied aggregate attention levels. An equilibrium attention vector satisfies:

$$\boldsymbol{M}(\boldsymbol{\chi}, \boldsymbol{\xi}) = \boldsymbol{g}(\boldsymbol{M}(\boldsymbol{\chi}, \boldsymbol{\xi}), \boldsymbol{\chi}, \boldsymbol{\xi}) \tag{45}$$

Here, the $\boldsymbol{g}$ mapping consists of the consumer g_c mapping and its counterpart for firms, which I denote g_f.

I use Gabaix's suggested formulations for the way in which firms and consumers approximate their losses from inattention. The derivations are simply a matter of matrix algebra, so I relegate them to Appendix A. To solve the model under indeterminacy, I follow Lubik and Schorfheide (2003, 2004). In the estimation section, I use their continuity solution to pin down a solution to the model. One could alternatively leave the solution unrestricted when estimating the model and estimate the additional parameters that yield the response of the endogenous variables to the fundamental shocks. However, as I show below, this leads to weak identification, and so is not my preferred approach.

There are two points worth noting about how the supply and demand shocks are motivated. I motivate demand shocks as stemming from fluctuations in the consumer discount factor rather than the level of technology, because Justiniano and Primiceri (2008) find discount factor shocks account for a large fraction of consumption variance.[13] As I note in Appendix B, the effect on consumer attention differs between the two shocks. The reason is that the discount factor shocks stem from within the consumer problem, and so affect their attention decision both direclty (by entering the equation for $\hat{c}_t$) and indirectly through the effect on the other endogenous variables. Technology shocks only affect the consumer problem through the indirect channel. As I explain in estimation section, however, whether one assumes technology or discount factor shocks is innocuous for the estimation results. The discount factor shock is generated by assuming that β fluctuates over time:

$$\beta_t = \beta + z_t$$
$$\tilde{z}_t = -\sigma\beta^{-2} z_t$$

The supply shock is motivated as a cost-push shock that represents all components of marginal cost other than real wages, such as commodity prices. Marginal cost is then:

$$\widehat{mc}_t = (\gamma + \phi)x_t + \omega^{-1}\eta_t$$
$$\omega := \frac{(1-\beta\theta)(1-\theta)}{\theta}$$

3.1 Multiple Regimes

In the section on average inflation targeting, I use a regime switching model. I assume there are two states: state N (for *N*ormal times), and state E (for *E*LB). I allow attention to vary with the regime; it

[13]See Figure 3.A in Justiniano and Primiceri (2008).

seems plausible that agents would choose a different level of attention in an ELB episode if the ELB leads to very different macroeconomic dynamics. To simplify the derviations, I assume state N is absorbing. Because it is an absorbing state, the choice of attention is not affected by switching probabilities, and so can be calculated as in the previous subsection. The state N model is as in the previous section, except for a different Taylor rule. I consider two monetary policy rules. The first is simply a Taylor rule without interest rate smoothing or monetary policy shocks:

$$i_t = \phi_\pi \pi_t + \phi_x x_t$$

The alternative is a rule in which the central bank responds to the average inflation rate over N quarters:

$$i_t = \phi_\pi \frac{1}{N} \sum_{s=0}^{N-1} \pi_{t-s} + \phi_x x_t$$

In state E, I assume that the economy faces an exogenous probability p of remaining in state E, but switches to state N with probability $1-p$. In state E, the Phillips and IS curve are given by:

$$\begin{aligned}
\pi_t &= \beta M_{f,E}(\boldsymbol{\chi}, \boldsymbol{\xi}) E_t \pi_{t+1} + \kappa x_t + \eta_t \\
x_t &= M_{c,E}(\boldsymbol{\chi}, \boldsymbol{\xi}) E_t x_{t+1} - \sigma(-E_t \pi_{t+1}) + \tilde{z}_t + \sigma r_E^n
\end{aligned}$$

Here, $r_E^n < 0$ is the natural real rate of interest in the ELB state. Note that I assume that this arises from a discount factor shock. In this case, the source of the shock does matter because it is an internal shock that affects consumer attention directly. In state E, the output gap, inflation, and real interest rate steady states differ from the steady state in normal times. In keeping with Gabaix's methodology, I assume that consumers discount future deviations from the normal times steady state.

The equations for calculating consumer and firm attention are now somewhat more complex, because the expectation at time t of the endogenous variables in time $t+h$ depends not just on whether the economy switched into state N but also *when* it did so; a switch in $t+1$ implies different $t+h$ state variables from a switch in $t+h$. I leave a description of the solution methodology to the Appendix.

4 Theoretical Applications

Sections 2 and 3, together with the associated appendices, extended Gabaix's formulation for the optimal choice of attention to a richer class of models, defined the notion of attention equilibrium, and discussed the existence of determinate and indeterminate equilibria. I also derived a refinement that shrank the region of the parameter space in which indeterminate equilibria exist. The refinement implies that, if sunspot shocks are small, where a determinate equilibrium exists, that will generally be a unique attention equilibrium. My first application looks at where indeterminate equilibrium does exist in the richer version of the model, and how this depends on the shock processes and the structural parameters. I then turn away from the determinacy condition, and show the importance of taking into account the endogeneity of the cognitive discount factor when assessing (i) changes in the Taylor rule parameters, and (ii) the effects of average inflation targeting.

4.1 Determinacy, Shocks, and Structural Parameters

The attainability refinement rules out indeterminate equilibria in some regions of the parameter space where the rational expectations Taylor principle fails. However, the determinacy condition is now considerably more complex than in the exogenous discounting case. The dependence of the determinacy condition on the endogenously chosen discount factors means that the determinacy region now depends strongly on the shock processes and structural parameters. In this section, I conduct some comparative statics to explore this further. I focus on variations in firm attention; such variations drive the results of the estimation in the empirical sections. Fluctuations in real variables, which determine consumer attention, are not large enough to increase m_c above its default even during the Great Inflation. I consequently leave discussion of the effects of the policy rule on consumer attention, which are somewhat more nuanced, to Appendix B.

In the calibrated examples, I set $\beta = 0.99$, $\phi = 1.0$, $\gamma = 2.0$, $\theta = 0.8$, $\rho_\eta = 0.75$, $\rho_z = 0.85$, $\rho_i = 0.6$, $\sigma_\eta^\varepsilon = 0.15$, $\sigma_z^\varepsilon = 0.35$, $\sigma_v^\varepsilon = 0.20$, and $\sigma_\zeta = 0.10$. For the high shock calibration, I adjust σ_η^ε up to 0.25. The low shock calibration sets σ_η^ε to 0.05. I use the continuity solution under indeterminacy. I set $m_{c,d} = m_{f,d} = 0.85$. Attention costs are $k_f = 1.5$, and $k_c = 4.5$. The rationale behind the attention parameter assumptions is discussed in section 6. I set $\phi_\pi = 0.60$ and $\phi_x = 0.20$.

If the economy is in the determinate region, an increase in ϕ_π generally reduces firm incentives to pay attention by reducing inflation volatility. The less volatile is inflation, the less firms change prices on average. As such, there is less of an incentive to pay attention to future deviations of endogenous variables, as this has a smaller effect on pricing. Hence, if m_f exceeds its default, and consumer attention is held fixed, an increase in ϕ_π reduces firm attention. The impact of ϕ_x on firm incentives to pay attention is ambiguous. For discount factor, or policy shocks, a higher ϕ_x stabilizes both the output gap and inflation, and so reduces firms' incentives to pay attention. For cost-push shocks, however, stabilization of inflation is achieved through inducing output gap volatility to offset the exogenous component of marginal costs. Stabilization of the output gap counteracts this, increasing firms' incentives to pay attention.

How does this affect determinacy? Figure 4.1 shows the determinacy region under (i) rational expectations, (ii) the case where attention is fixed at its default level, and (iii) endogenous attention, using the same baseline calibration as before.[14] Under endogenous attention, the boundary of the determinacy region lies below the boundary under rational expectations. However, it lies above the boundary in a model with fixed attention. As ϕ_π decreases, M_f rises above its default and pushes the economy into indeterminacy. As such, monetary policy must react more strongly to inflation than in the fixed attention model to ensure determinacy. This illustrates how the determinacy condition in a fixed attention model is vulnerable to the Lucas critique. Note that ϕ_x has a larger effect on determinacy in the behavioural model than under rational expectations; equation (41) illustrates that the lower is M_f, the more important is ϕ_x for the determinacy condition.

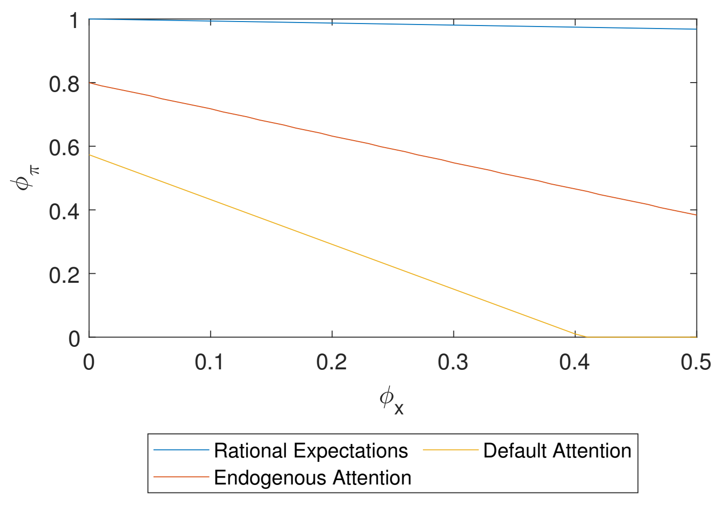
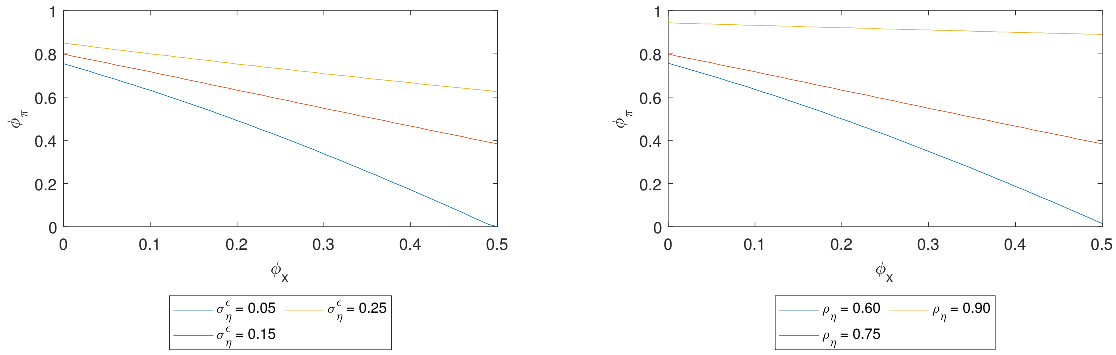


Figure 4.1: Determinacy regions for the baseline calibration.

Incentives to pay attention are strictly increasing in the variances of fundamental shocks. As such, if m_f exceeds its default, a rise in these parameters increases firm attention and shrinks the determinacy region. Under the baseline calibration, σ_η^ε is set to 0.15. Figure 4.2a shows how the determinacy region expands if σ_η^ε decreases to 0.05, and contracts if σ_η^ε rises to 0.25. Incentives to pay attention are also increasing in the shock autocorrelation parameters. In the extreme case where the persistence of each

[14] As noted above, the attainability refinement means there is only a small region where both determinate and attainable indeterminate equilibria exist. As such, I just show the boundary of the region that allows for determinate equilibrium.

shock is zero, there is no incentive to pay attention to future deviations of the endogenous variables; whether the agent pays attention or not, deviations are zero in expectation for all future time periods. In the baseline calibration, ρ_η is 0.75. Figure 4.2b shows that if ρ_η falls to 0.60 then attention decreases and the determinacy region widens. Notice that high persistence ($\rho_\eta = 0.9$) leads to very high attention; the determinacy region is then little changed from the rational expectations case.

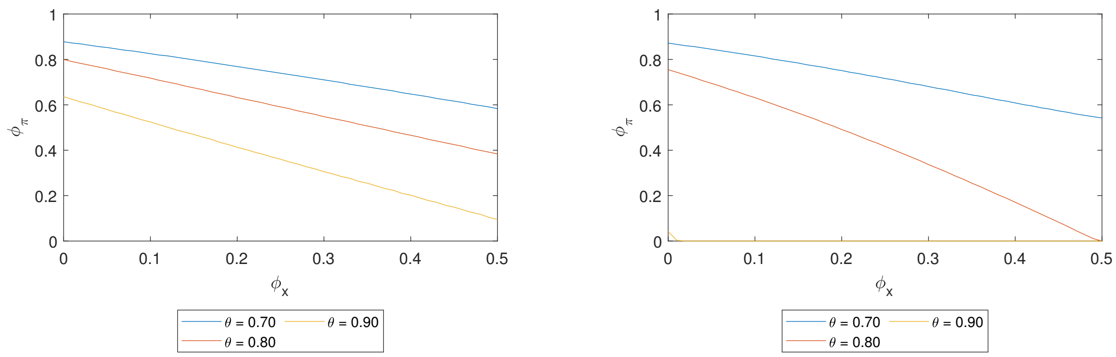


(a) Varying σ^ε_η (b) Varying ρ_η

Figure 4.2: Determinacy region with changing cost-push shock variance and persistence.

Equation (41) shows that θ and γ are more important for the determinacy condition under endogenous attention than rational expectations. First, these parameters determine κ and σ, which have a larger effect on the determinacy condition when M_f and M_c are less than unity. Second, these parameters affect the attention problems of firms and consumers. If θ is higher, firms expect prices to persist for longer, so incentives to pay attention to the future are greater. If γ is higher, consumers have a stronger preference for consumption smoothing, so paying attention affects their decisions less.

In numerical examples, θ has a particularly powerful effect on the determinacy region, because small changes in θ induce large changes in κ. This effect is illustrated in figure 4.3. In the baseline calibration, θ is 0.80. In each figure, I show the effect of changing θ to 0.7 or 0.9. In the 4.3a, the baseline shock calibration is used. Generally, a higher θ leads to a flatter Phillips curve slope, expanding the determinacy region. This may be partially offset by the greater firm attention induced by longer price duration. In 4.3b, I use the low shock calibration, where attention is close to its default. When attention is low, changes in θ may have larger effects, because of the interaction between κ and $(1 - \beta M_f)$ in (41). Notice how large is the effect of changing θ to 0.90; the indeterminacy region almost disappears.



(a) Baseline shock variance (b) Low shock variance

Figure 4.3: Determinacy region with changing Calvo pricing parameter

4.2 Changes in the Taylor Rule and Macroeconomic Volatility

I now focus on policy rules which ensure determinacy, and assess how changes in the policy rule affect macroeconomic volatility. As above, I focus on the effects of changes in firm attention. I use the same calibration as for the comparative statics exercise above.

As in rational expectations models, policy has a direct effect on volatility, because ϕ_π determines the response of inflation and the output gap to shocks. There is also, however, an indirect effect, because ϕ_π affects m_f, which in turn affects the response of the endogenous variables to shocks. Figure 4.4a shows how equilibrium firm attention changes as ϕ_π rises from 1.0 to 2.5. Notice how $g_f(\boldsymbol{M}, \boldsymbol{\chi}, \boldsymbol{\xi_f})$ shifts down under the higher ϕ_π; the red line showing $g_f(\boldsymbol{M}, \boldsymbol{\chi}, \boldsymbol{\xi_f})$ for $\phi_\pi = 2.5$ is always at the default. Figure 4.5a shows the impact of this on inflation volatility. The blue dots show the standard deviation of inflation for different levels of ϕ_π under endogenous attention. The red dots show the same series but fixing attention at the default. At $\phi_\pi = 1.0$, volatility is higher under endogenous attention as attention exceeds its default. At $\phi_\pi = 2.5$, volatility is the same in either circumstance, as under endogenous attention m_f is at its default. As such, the effect of changing ϕ_π from 1.0 to 2.5 on inflation volatility is larger under endogenous attention. How far changes in attention affect inflation volatility depends on the parameterisation. I explore this in more detail in Appendix B.

Equally importantly, irrespective of the change induced in m_f by a change in ϕ_π, the higher attention induced by larger shocks means that the effects of changing ϕ_π are larger. The reason is that ϕ_π has a role in dampening expected inflation, which matters more for inflation today if firms are paying more attention to the future. When shocks are large and ϕ_π is low, attention rises above its default, raising the costs associated with a weak policy response to inflation. Once again, this channel is absent from a model in which the level of discounting is exogenous and does not depend on the size of the shocks hitting the economy.

An important feature of the sparsity-based framework is that these behavioural effects only emerge when shock are large and persistent. Otherwise, attention is anchored at its default level and increases in ϕ_π do not affect attention. The impact of the policy rule on inflation volatility consequently tends to be greater when shocks are large. Figure 4.4b shows the same policy experiment as 4.4a but with lower shock variances. Here, whether ϕ_π is set at 2.5 or 1.0, attention remains at its default (hence only the red line is visible). As such, there is no behavioural monetary policy channel. In figure 4.5, then, inflation volatility is the same under endogenous or fixed attention (so again only the red series is visible).

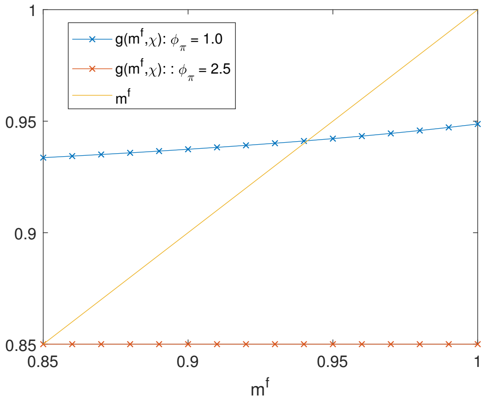
Although m_f is decreasing in ϕ_π when it exceeds the default level, how large is the impact of a given change in ϕ_π on m_f depends on the parameterisation. In particular, if θ is very high, then large changes in ϕ_π are required to have a meaningful effect on m_f. Recall that marginal cost is given by:

$$\widehat{mc}_t = (\phi + \gamma)x_t + \omega^{-1}\eta_t$$

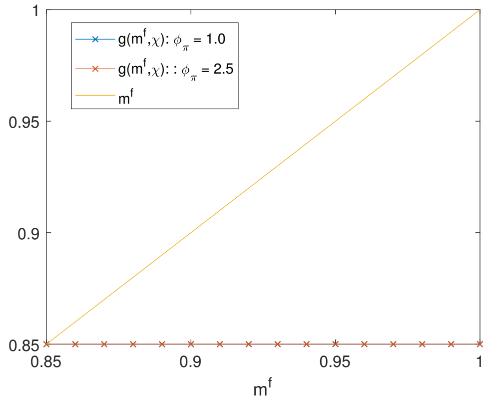
As before, ω is the Phillips curve slope in terms of marginal costs. When θ is high, ω is low, so marginal costs are more strongly driven by the exogenous shock rather than the output gap. As such, a stronger monetary policy response to inflation is required to cause a meaningful reduction in marginal cost volatility and hence firms' incentives to pay attention to the future.

4.3 Average Inflation Targeting

Average inflation targeting may reduce the costs of an occasionally binding ELB on interest rates under rational expectations. However, as Bundiato et al. (2021) show, average inflation targeting may have small effects under exogenous cognitive discounting. This is because average inflation targeting operates through altering expectations while the economy is in the zero lower bound state, through affecting the inflation and output that would be realized if the constraint ceases to bind in future.

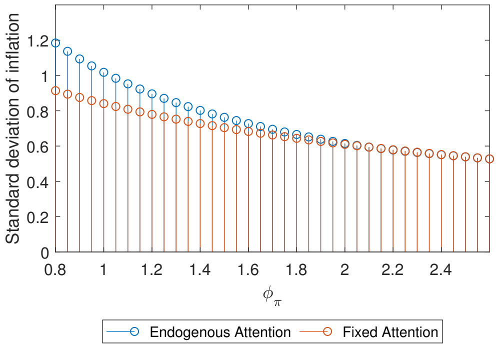


(a) High Shock Calibration

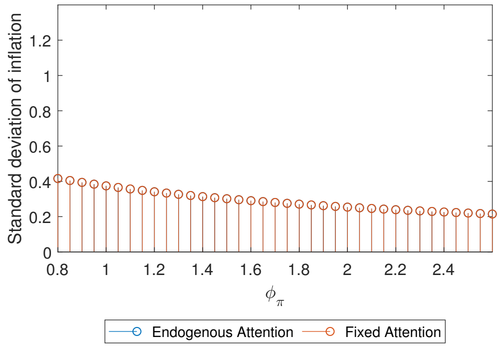


(b) Low Shock Calibration

Figure 4.4: Equilibrium attention for different values of ϕ_π



(a) High Shock Calibration



(b) Low Shock Calibration

Figure 4.5: Standard deviation of inflation for different levels of ϕ_π

I now show that the effects are potentially far larger under endogenous cognitive discounting, using the model described in section 3. The calibration considered is: $\beta = 0.99$ $\theta = 0.85$, $\gamma = 3$, $\phi = 1$, $\phi_\pi = 2.50$, $\phi_y = 0.25$, $\rho_\eta = 0.60$, $\rho_z = 0.80$, $\sigma_{e,\eta} = 0.10$, and $\sigma_{e,z} = 0.20$. The strong monetary policy response to inflation gives AIT the greatest chance to affect macroeconomic dynamics. I assume $p = 0.85$, giving an expected duration of the ELB episode of 6 to 7 quarters. I assume that the natural real rate is -1.5% during the period of the ELB. I use a slightly lower default level of attention than in the previous section, $m_{c,d} = m_{f,d} = 0.70$, since Bundiato et al. (2022) emphasise that the power of AIT is muted particularly when the discount factor is far from rationality. I continue to assume $\kappa_c = 4.5$ and $\kappa_f = 1.5$. I then consider two versions of the policy rule. The first is that the central bank follows a simple Taylor rule in state N. The second assumes that the central bank responds to the average inflation rate over an extended period. I choose a long window of 16 quarters, again giving AIT the best chance of mitigating the costs of the ELB.

Because state N is absorbing, one can use the following steps to solve the model. First, one has to find the equilibrium level of m_c and m_f in state N. In the calibration chosen, the equilibrium in all cases involves $m_{c,N} = m_{c,d}$ and $m_{f,N} = m_{f,d}$. One can then solve for the state N dynamics, which in turn allows one to solve for the state E dynamics for any combination of $m_{c,E}$ and $m_{f,E}$. Then, one can use standard numerical routines to find fixed points of the attention vector $\boldsymbol{m_E}$.

Figures 4.6 and 4.7 show the implied optimal choices of $m_{f,E}$ and $m_{c,E}$ for different combinations of $m_{c,E}$ and $m_{f,E}$ under AIT and under the Taylor rule. In the present calibration, high attention can push the economy into the indeterminate region when monetary policy is set using a Taylor rule. At the boundary of the determinacy region, the recession induced by the ELB becomes unboundedly large.

That generates the dramatic increase in the choice of attention close to the boundary; the choices of $m_{c,E}$ and $m_{f,E}$ approach unity at the boundary. I have not yet investigated the possibility of equilibria in the indeterminate region, but this is an important endeavour for future research; recall from the discussion of determinacy above that determinate equilibria are not guaranteed to exist when policy is passive.

Figures 4.8 show equilibria in $m_{f,E}$ and $m_{c,E}$ under the two monetary policy regimes. The red lines show which combinations of $m_{f,E}$ and $m_{c,E}$ imply that the same $m_{c,E}$ is the optimal choice of attention for consumers. The blue lines show the same for $m_{f,E}$. Attention equilibria occur where the two lines intersect; the intersections show points where a particular combination of $m_{f,E}$ and $m_{c,E}$ imply macroeconomic dynamics that justify same combination of $m_{f,E}$ and $m_{c,E}$ as optimal choices. Under average inflation targeting, the unique attention equilibrium is with $m_{c,E} = m_{c,d}$ and $m_{f,E} = m_{f,d}$. The analysis is simplified considerably in this case because even full attention does not push the economy into the indeterminate region in this calibration.

Under the Taylor rule, the situation is more complex. A first point to note is that the optimal choice of firm attention is much higher. Under AIT, the expected rate of inflation is a combination of (i) the low rate expected if the economy remains in state E, and (ii) the high rate that will be justified over an extended period if the economy moves into state N. Expectations are consequently closer to zero, and incentives to pay attention are lower. Under the Taylor rule, effect (ii) is absent, and so expectations are further from zero, and incentives to pay attention are much higher.

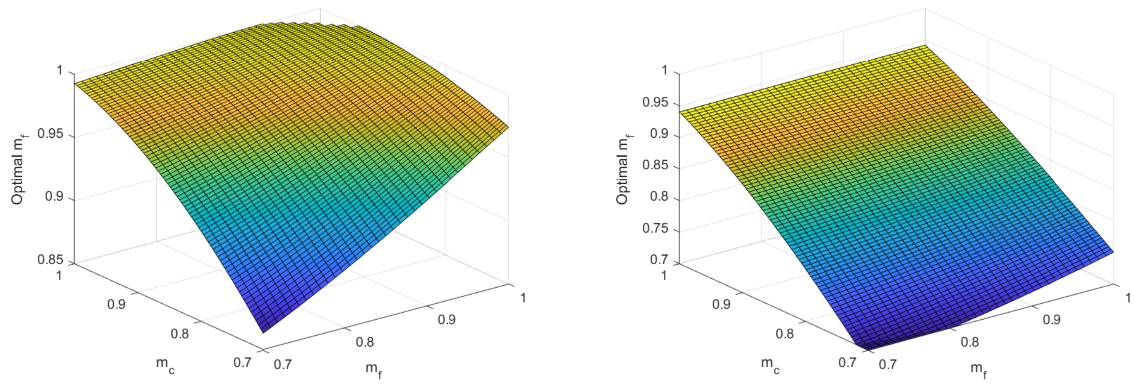
The second point to notice is that for any given $m_{f,E}$ there may be two or even three equilibria in $m_{c,E}$. If consumers pay little attention, then output and inflation are not too far from zero. The real interest rate consequently does not rise too much, and consumers have low incentives to pay attention. However, if consumers pay sufficient attention, then the economy approaches the boundary of the determinacy region, an arbitrarily large recession and deflation ensues, the real interest rate rises dramatically, and incentives to pay attention are high.

This effect results in two attention equilibria existing under the Taylor rule. These are summarized in table 1, which shows the equilibrium levels of attention under AIT and the two Taylor rule equilibria, the resulting average deviation in inflation and the output gap in state E, and the variability of inflation and the output gap in the two states. Note in particular that the average deviation of inflation from the target is dramatically higher under the Taylor rule, and the Taylor rule also induces a considerably larger recession. To show the source of this difference, I analyze three counterfactuals. First, I show the dynamics if one keeps average inflation targeting but uses the equilibrium attention level under the two Taylor rule equilibria. Note the size of the deflation and recession induced by the ELB is much smaller than under the Taylor rule; this is because AIT is particularly strongly effective when agents are close to rationality and hence very forward-looking. I then analyze a counterfactual where the policy rule used is the Taylor rule but the attention levels are the AIT attention levels. Here, the size of the deflation and recession induced by the ELB is quite similar to the AIT case.

To summarize, AIT is effective at reducing the costs of the ELB when agents are very forward-looking, and less so when agents discount the future heavily. However, average inflation targeting induces agents to discount the future far more than does the Taylor Rule. As such, the benefits of AIT may be higher under endogenous than exogenous discounting. The possibility of multiplicity does, however, complicate the analysis, and the extent of the differences in equilibrium attention levels between the two regimes likely depend on the calibration. Future work could explore this further.

5 Empirical Analysis: Identification

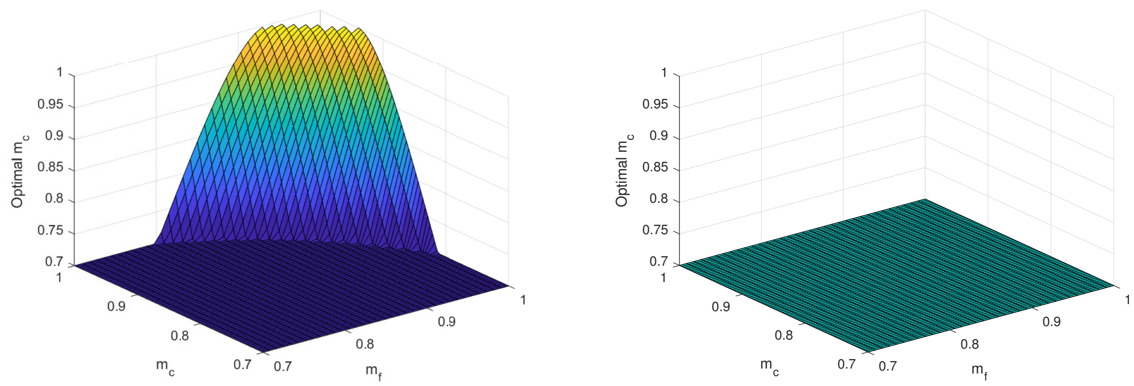
I now argue that endogenizing attention helps to avoid weak identification. Gabaix (2020) shows that in a single equation example with an unobserved forcing variable attention parameters may not be identified,



(a) Taylor Rule

(b) Average Inflation Targeting

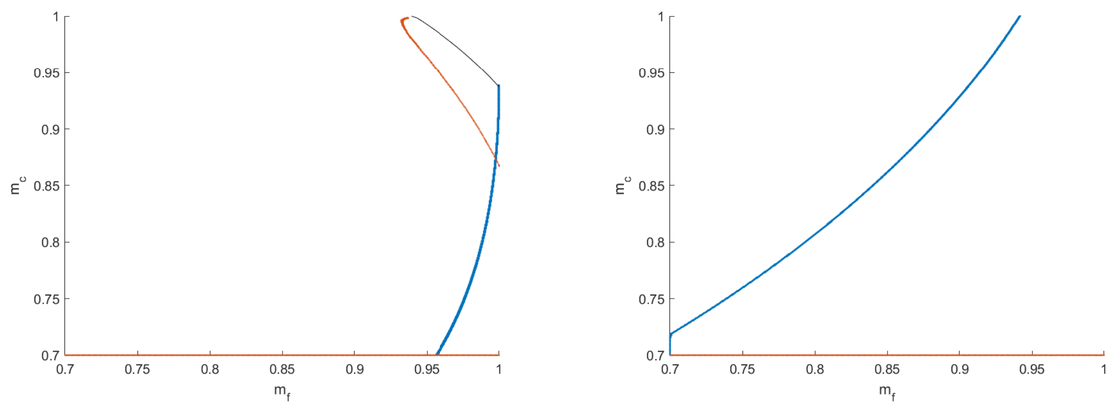
Figure 4.6: Optimal choice of $m_{f,E}$



(a) Taylor Rule

(b) Average Inflation Targeting

Figure 4.7: Optimal choice of $m_{c,E}$



(a) Taylor Rule

(b) Average Inflation Targeting

Figure 4.8: Attention Equilibria. Note that points marked in red are equilibria in $m_{c,E}$, and blue denotes equilibria in $m_{f,E}$. Intersections of the blue and red curves give attention equilibria. The black line denotes the boundary of the determinacy region.

Table 1: Equilibria under Taylor Rule vs. Average Inflation Targeting

	Avg. Inflation Targeting			Taylor Rule		
	Equilibrium	Counterfactual 1	Counterfactual 1	Equilibrium 1	Equilibrium 2	Counterfactual
$m_{c,E}$	0.70	0.70	0.87	0.70	0.87	0.70
$m_{f,E}$	0.70	0.96	1.00	0.96	1.00	0.70
$\bar{\pi}_E$	-0.31	-0.54	-0.91	-1.18	-5.80	-0.42
$\bar{y}_E$	-1.28	-1.31	-1.95	-2.06	-8.32	-1.53
$St.Dev.\pi_N$	0.64	0.64	0.64	0.59	0.59	0.59
$St.Dev.y_N$	0.64	0.64	0.64	0.64	0.64	0.64
$St.Dev.\pi_E$	0.73	1.01	1.01	1.28	1.47	0.78
$St.Dev.y_E$	0.29	0.37	0.46	0.50	0.72	0.29

because they have the same effect on the observed moments as the volatility of the forcing variable.[15] I show that a similar logic causes weak identification in a systems of equations approach. Andrade et al. (2019) also note the possibility of weak identification. They note that this arises from the shock autocorrelations being similar, an observation due to Andrews and Mikusheva (2014). My analysis of identification differs from theirs in that (i) I consider a setting with three observables rather than two,[16] (ii) I assume firms and consumers have different cognitive discount factors, and (iii) I also consider endogenous attention.[17]

I first assess point identification, following Iskrev (2010a).[18] I focus on the model with interest rate smoothing, and on the determinate case.[19] Under endogenous attention, the econometrician estimates:

$$\boldsymbol{\chi} = \begin{bmatrix}\phi_\pi & \phi_x & \theta & \gamma & \rho_\eta & \rho_z & \rho_i & \sigma_\eta^e & \sigma_z^e & \sigma_v^e\end{bmatrix}$$

The econometrician is assumed to know the structure of the attention problem, and thus can calculate the implied M_f and M_c for any parameter vector. Under exogenous attention, the econometrician must estimate M_f and M_c.[20] I find that:

- Under exogenous attention, identification fails if ρ_z is equal to ρ_η or ρ_i.
- Under endogenous attention, identification does not fail in either of these circumstances.

Points of identification failure may be surrounded by wide regions of weak identification.[21] Additionally, even in areas of the parameter space far from points of identification failure, I find that the difficulty in distinguishing attention parameters from shock volatilities leads to weak identification. I show this using a method due to Andrews and Mikusheva (2014). They note that under strong identification both the quadratic variation of the score vector and the negative of the Hessian of the log likelihood converge to the theoretical Fisher information matrix. Under weak identification, only the former does so. One

[15]See Gabaix (2020) p.15.

[16]In their section 3 where they analyse identification, Andrade et al. (2019) assume only the output gap and inflation are observable, although their estimated model features three observables.

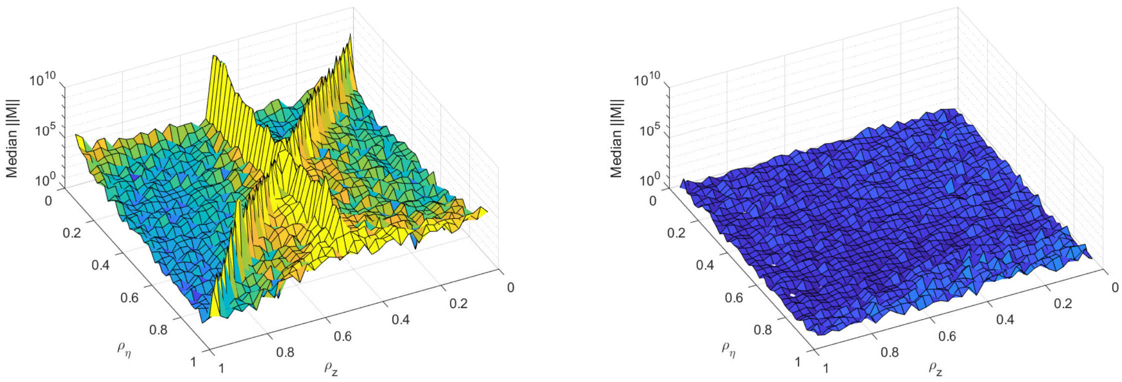
[17]In the Andrade et al. (2019) example, when their two autocorrelation parameters are equal, they have two degrees of under-identification even after imposing assumptions on β, θ, ϕ, and the policy parameters. Hence, knowing the cognitive discount parameter (which in their case is the same for firms and consumers) would still result in identification failure at this point absent further assumptions. Hence I believe that my conclusion is specific to the three observable case.

[18]This method involves calculating a vector of the first and second moments analytically. As shocks are Gaussian, the first two moments completely characterise the restrictions imposed by the model. One then calculates the Jacobian of this vector with respect to the parameters. If the Jacobian is full rank at a given parameter vector, then each parameter has a distinct effect on the first two moments, and so the model is point identified.

[19]See Appendix C for details of identification in the model without interest rate smoothing.

[20]Note that several papers in the literature, such as Andrade et al. (2019), impose the restriction that firms and consumers have the same cognitive discount factor, which might help with identification. However, I note that the microfoundation shows that these cognitive discount factors are driven by rather different factors; for the consumer problem, real interest rate volatility is the key driver, whereas for firms it is marginal cost and inflation volatility. Indeed, my results suggest that the two should differ very substantially.

[21]This point is noted by Andrews and Mikusheva (2014). In their example, weak identification arises even if the difference between autocorrelation parameters is as large as 0.3.



(a) Exogenous Attention

(b) Endogenous Attention

Figure 5.1: Strength of Identification under Exogenous and Endogenous Attention.

can then assess the strength of identification using the scaled difference between the two:

$$M = J_T^{-\frac{1}{2}}(I_T - J_T)J_T^{-\frac{1}{2}}$$

Here J_T is the quadratic variation of the score vector, and I_T is the negative Hessian. If identification is weak, the largest eigenvalue of M in absolute value, denoted $||M||$, becomes large. I use the baseline calibration, but with ϕ_π set to 1.5 to ensure determinacy. I simulate samples of 100 quarters for a range of different values of the shock autocorrelations ρ_z and ρ_η. Recall that ρ_i is set to 0.6. The median $||M||$ under exogenous and endogenous attention is shown in figure 5.1; this measure becomes exceptionally large at many points in the parameter space under exogenous attention. Under exogenous attention, not only is the median $||M||$ exceptionally high around the points of identification failure, but even far from these points it remains large;. even at the points of strongest identification it is c.40. Under endognenous attention, the measure is far smaller (often c.2), indicating far stronger identification.

To assess the source of this issue, I find the correlation matrix of the score vector as follows:[22]

$$\tilde{J}_T = D^{-\frac{1}{2}} J_T D^{-\frac{1}{2}}$$
$$D = diag(J_T)$$

The correlations relating to M_f and σ_η^ε are often as high as 0.999. This suggests the source of weak identification is similar to the reason for identification failure Gabaix's single equation example; attention parameters and shock volatilities have similar effects on the model's second moments.

The application to the Great Inflation and Great Moderation uses a Bayesian approach. If the likelihood is weakly identified, then the posterior is determined largely by the priors. Under exogenous attention it is difficult to ascertain whether, for example, a high variance of inflation results from high attention or volatility in the cost-push shock; there is a range of combinations of M_f and σ_η^ε that match the data. Which combination is more likely per the posterior depends largely on the priors. Under endogenous attention, a combination of M_f and σ_η^ε is selected which is consistent with optimising behaviour.[23]

An alternative approach is to use methods robust to weak identification. Andrade et al. (2019) conduct both a single equation GMM estimation of the IS and Phillips curves in a behavioural model,

[22] This decomposition is proposed by Iskrev (2010b) for the theoretical information matrix; see equation (3.3) in Iskrev (2010b). I use it here on J_T. As per Andrews and Mikusheva (2014), J_T converges to the theoretical information matrix.

[23] For procedures for assessing identification specific to a Bayesian context, see Koop et al. (2013). Note that simply comparing priors and posteriors is not sufficient to check identification. As Koop et al. argue, requiring the parameters to be consistent with determinacy or indeterminacy imposes a joint restriction that may lead posteriors to differ from priors even if they are unidentified. This matters because I estimate the model imposing either determinacy or indeterminacy.

and a system of equations maximum likelihood estimation. They use methods proposed by Andrews and Mikusheva (2015) and Andrews (2018) to generate robust confidence sets for the attention parameters. However, their robust confidence intervals are very wide. In their GMM estimation, for example, the interval for M_f ranges from 0.14 to 0.95. Values at the lower and upper ends of this interval have vastly different implications for indeterminacy. Moreover, large regions within this intervals are inconsistent with the model's microfoundations. Endogenising attention avoids this issue by finding a combination of attention and shock volatilities consistent with both the data and the microfoundations.

However, identification under endogenous attention relies on correct specification of the attention problem. One must make assumptions on the form of the attention cost, the default level of attention and the attention cost. If one estimates these parameters, attention can vary independently of the shock variances, which is the issue that affects the exogenous case. There are as yet few estimates for these attention costs. As such, any empirical results obtained are necessarily tentative. For my application to the Great Inflation and Great Moderation, I argue for a particular set of attention costs as plausible, but also test a range of values to check the robustness of the results. I then present a method of externally validating the results using empirical evidence on expectations data.

6 Empirical Applications

The last section demonstrated that endogenizing the cognitive discount factors can overcome a weak identification problem suffered by exogenous discounting models. I now demonstrate two applications of this method. First, I show that with plausibly calibrated attention costs, the endogenous attention model rules out indeterminacy as a cause of the Great Inflation. Second, I show that the endogenous attention model implies that the slopes of the Phillips and IS curves would appear to flatten during times of macroeconomic stability. The correct slopes can be accurately recovered by estimating an endogenous discounting model, whereas the exogenous discounting model gives imprecise estimates.

6.1 Great Inflation and Great Moderation

I now estimate the model on the Great Inflation and Great Moderation periods. This allows me to show empirical relevance of endogenous discounting for identification, determinacy, and the effects of policy.

6.1.1 Model

I estimate the model given by (36), (37), and (38). As in section 5, I assume that β and ϕ are known, and take values of 0.99 and 1.0 respectively. Attention is considered endogenous, and I once again assume values for default attention and the attention costs. Under determinacy, I estimate the parameters:

$$\boldsymbol{\chi} = \begin{bmatrix} \phi_\pi & \phi_x & \theta & \gamma & \rho_\eta & \rho_z & \rho_i & \sigma_\eta^\varepsilon & \sigma_z^\varepsilon & \sigma_v^\varepsilon \end{bmatrix} \tag{46}$$

Under indeterminacy, I impose the the continuity solution, to avoid issues of weak identification discussed in section 5.[24] I then estimate the parameter vector above, augmented with σ_ζ.

As in section 5, I specify default levels of attention and attention costs. The attention parameters have a considerable impact on the results. By specifying an arbitrarily high cost of attention, and a low default level, one could ensure attention is sufficiently low such that the model is always determinate, resulting in a arbitrarily small marginal likelihood for the indeterminate model. Equally, if one sets the cost of attention arbitrarily low, the model reduces to the rational expectations model, making indeterminacy very likely in the Great Inflation sub-sample because the interest rate response to inflation is weak.

I calibrate default attention to $m_{f,d} = m_{c,d} = 0.85$. This corresponds to the attention level M_c suggested by Gabaix (2020) for calibrating the behavioural new Keynesian model. As Gabaix notes, this

[24]Nonetheless, I note that this may disadvantage the indeterminate model in the model comparison conducted below. Future research could assess the sensitivity of the results to allowing response to fundamental shocks to vary.

implies discounting of c.50% at a one-year horizon. Gabaix notes that this value is broadly in line with studies estimating the weight on forward looking terms in the Phillips and IS curves, such as Gali and Gertler (1999) or Fuhrer and Rudebusch (2004).

For consumers, I assume an attention cost of $k_c = 4.5$, which is the value estimated by Ganong and Noel (2017).[25] To my knowledge, this is the only estimate for scale-free attention costs; no estimates exist for firms. One would expect costs for firms to be lower. My baseline specification assumes $k_f = 1.5$. I calculate that if paying attention to a variable changes pricing decisions by 1.5% on average, then doing so increases profits by c.0.1% of sales. This accords with Coibion, Gorodnichenko, and Ropele's (2019) finding that firms lose profits of roughly c.$0.1 - 0.3$% of sales as a result of inaccurate inflation perceptions. For details of my calculation, see Appendix D. To show the robustness of the results to this assumption, I also show an alternative specification in which k_f is half the baseline level at 0.75.

I impose the requirement that the equilibrium found must be attainable. When calculating the unconditional covariance matrix for the variables, which I use to initialize the Kalman filter, I suppose that the first period of the sample is the first period that the economy was in the indeterminate state. In theory, one should allow attention to drift upwards over time, as explained in section 5. In practice, it is then technically challenging to calculate the likelihood. I instead hold the level of attention fixed over time in the indeterminate state. Moreover, I assume that if an indeterminate equilibrium exists in the first period, then it exists in all subsequent periods.

One has to make an assumption about the source of IS curve shocks to calculate consumer attention. Justiniano and Primiceri (2008) find that discount factor shocks account for a large part of consumption variance. For simplicity, I therefore assume all IS curve shocks are driven by the discount factor.

6.1.2 Data

All data are from Federal Reserve Economic Data. The baseline specification follows Coibion and Gorodnichenko (2015b) in using unemployment (UNRATE) as the forcing variable in the Phillips curve. I filter the unemployment rate using the Hodrick-Prescott (HP) filter to obtain the unemployment gap, and use the negative of this as an output gap measure.[26] As a robustness check, I follow Lubik and Schorfheide (2004) in using the HP filter to de-trend the log of real GDP per capita (A939RX0Q048SBEA). Deviations from trend are multiplied by 100 to obtain a percentage point output gap measure. I include this specification because one might expect the choice of forcing variable to affect the Phillips curve slope and hence the estimated θ. As noted above, θ considerably affects the determinacy region.

For inflation, I use the quarter-on-quarter percentage change in the GDP deflator (GDPDEF). For the nominal interest rate I use the effective Federal Funds rate (FEDFUNDS), which I convert to a quarterly rate. I de-mean the nominal variables in each sub-sample.[27] I prefer to de-mean nominal variables, rather than HP filtering. Canova and Ferroni (2011) note that "while real variables typically show long run drifts, nominal variables just display low frequency fluctuations".[28] I find removing such fluctuations a concern when estimating the model under indeterminacy. As highlighted in section 2, sunspot shocks can have very persistent effects on inflation; de-trending could remove this low frequency variation.

For the Great Inflation period, I use data from 1960:I to 1979:II, and for the Great Moderation period I use 1984:I to 2007:IV. I also report results from 1990:I to 2007:IV, motivated by Fernandez-Villaverde et al.'s (2010) finding that the policy response to inflation in this period was weak.

[25] See p.24 of Ganong and Noel (2017). In their model, they interpret this parameter as "the largest possible income shock for which the agent would not cut spending in advance at all".

[26] For the HP filter, I set $\lambda = 1600$, as standard in the literature. To avoid the end point problems associated with the HP filter, noted by St Amant and van Norden (1997), I de-trend a longer time series and take sub-samples of the longer series. As such, the first and last four observations of the filtered data are not used in any sub-sample.

[27] Lubik and Schorfheide instead include additional parameters for their average levels.

[28] See Canova and Ferroni (2011), p.74.

6.1.3 Priors

Table 2: Priors

Parameter	Distribution	Mean	St. Dev.	90 pct. interval
ϕ_π	Γ	1.10	0.50	[0.43,2.03]
ϕ_x	Γ	0.25	0.15	[0.06,0.54]
θ	B	0.70	0.075	[0.57,0.82]
γ	Γ	2.00	0.50	[1.25,2.89]
ρ_η	B	0.70	0.10	[0.52,0.85]
ρ_z	B	0.70	0.10	[0.52,0.85]
ρ_i	B	0.50	0.20	[0.17,0.83]
σ_η^ε	Γ^{-1}	0.30	0.50	[0.08,0.80]
σ_z^ε	Γ^{-1}	1.00	1.00	[0.32,2.43]
σ_v^ε	Γ^{-1}	0.30	0.50	[0.08,0.80]
σ_ζ	Γ^{-1}	0.30	0.50	[0.08,0.80]

For the policy parameters, shock autocorrelations, and the coefficient of relative risk aversion, I set the mean and standard deviations for each prior in line with Lubik and Schorfheide (2004).[29] Lubik and Schorfheide estimate the Phillips curve slope directly, so I cannot follow their prior for θ. I use a beta distribution to constrain θ to the interval (0,1). I centre the distribution at 0.7, giving an average price duration of 10 months, close to the average duration found in Nakamura and Steinsson (2008). The 90% credibility interval implies price duration between 7 months and 17 months.[30] For shock volatilities, I use similar means to Lubik and Schorfheide, but larger standard deviations; these parameters matter for attention and hence determinacy, so I prefer to give more weight to the data.

6.1.4 Methodology

My process for estimating the model broadly follows techniques outlined in Lubik and Schorfheide (2004) and Herbst and Schorfheide (2015), adapting these methods to allow for endogenous attention. The key difference is that for each set of parameters drawn, the equilibrium level of attention must be calculated in order to evaluate the log-likelihood. A full description of the algorithm is given in Appendix E.

6.1.5 Results

Table 3 compares the log marginal likelihood of determinacy and indeterminacy for each sub-sample in each specification, and shows the evidence against indeterminacy according to Kass and Raftery's (1995) descriptive scale. Tables 4 shows full results for the specification. Full results for the alternative specifications are shown in Appendix E.

Comparison of the marginal likelihood suggests that, for the Great Inflation period, the evidence in favour of determinacy is "'very strong" in the baseline specification, despite the mean of ϕ_π lying well below unity. This result also holds in the specification using the output gap. Even in the low cost of attention specification, the evidence against indeterminacy is "strong" on the Kass and Raftery scale. As such, I believe that the findings represent more conclusive evidence against the good policy hypothesis than is found in the exogenous attention estimation of Ilabaca et al. (2020).

Why does this result emerge? First, even under endogenous attention, the estimated Phillips curve is rather low. As in section 4, this generally pushes the economy towards indeterminacy. Second, for consumers, the Ganong and Noel (2017) estimate of k_c implies that in the present specification consumer attention always remains at its default level. In this model consumers care about real interest rates. Fluctuations in the real rate are smaller than nominal fluctuations in this period, because ϕ_π is close to the autocorrelation parameter for the cost-push and demand shocks. With a fairly high consumer cost of

[29]My prior for γ corresponds to Lubik and Schorfheide's prior for the inverse of the IS curve slope.

[30]I calculate credibility intervals using the 5th and 95th percentiles of the distribution. Note that this differs from Lubik and Schorfheide (2004), who calculate the shortest such interval.

attention, the volatility in the real rate is insufficient to move attention above the default level. It might be that a richer model with medium-scale features might cause consumers to pay more attention. Third, as noted in section 4, ϕ_x matters more for determinacy under endogenous attention than under rational expectations. In the unemployment gap specification, ϕ_x tends to be rather high.

As such, one would require an exceptionally high m_f to generate indeterminacy. The results do suggest that m_f was well above its default level during the Great Inflation period in response to the high variance and persistence of cost-push shocks. However, firms remain some way from rational expectations. As I discuss in section 7, this pattern is externally validated by empirical expectations data from Coibion and Gorodnichenko (2015a).

The reader might wonder why the marginal likelihoods of the determinate and indeterminate models are so different, despite the estimates for many parameters seeming to be close. However, the estimates imply rather different inflation volatility and persistence. This can be seen from the high m_f obtained under indeterminacy. Indeed, the model is pushed towards generating high inflation volatility and persistence because m_f must be very high for system to be indeterminate. The reason is the difference in ρ_η. When ϕ_π is low, small differences in ρ_η dramatically change the response of inflation to cost-push shocks, and this more than offsets the lower estimate for σ_η^ε under indeterminacy.

For the Great Moderation, indeterminacy is strongly rejected in all specifications. For the 1984:I to 2007:IV sub-sample, I obtain a mean estimate for ϕ_π above unity in all specifications, in line with the literature. Like Fernandez-Villaverde et al. (2010), however, I find that the point estimate for ϕ_π falls below unity for the period after 1990, albeit the 90% credibility interval is very wide. This result holds across all specifications. Despite this, the probability of indeterminacy is negligible even in the lowest attention cost specification. Indeterminacy is particularly unlikely for this period because (i) the cost-push shock is less volatile and persistent than in the Great Inflation period, lowering firm attention, and (ii) the estimate for θ is even greater. Notice how extreme the estimate must be (ϕ_π of 0.20) to generate indeterminacy; this leads to the very low marginal likelihoood. I note that Ilabaca et al. (2020) find a neglibible probability of indeterminacy in their Great Moderation sub-sample (which beings in 1982:I), but they do estimate a high ϕ_π of 2.28 under determinacy during this period. Under endogenous attention, indeterminacy is very unlikely for this period even in a sub-sample where ϕ_π appears relatively low.

Note that this model assumes that persistence in inflation arises from the persistence of cost-push shocks, and the results outlined above show that this matters for determinacy. Other approaches assume that this arises from indexation.[31] Benati (2008) finds that the extent of indexation has changed substantially over time. Future research might model the degree of backward looking behaviour endogenously as a behavioural phenomenon, perhaps using similar techniques to those outlined in this paper.

6.1.6 Counterfactuals

I now consider how the economy would have behaved under counterfactual policy rules. I begin by considering determinacy regions. Figure 6.1 shows the determinacy regions generated by the structural and shock process parameters for the Great Inflation and post-1990 Great Moderation periods under the determinate estimate in the baseline specification.[32] Figure 6.2 shows results for the low cost of attention specification. I use the mean as a point estimate.[33]

Although the estimated rule for the Great Inflation leads to indeterminacy under rational expectations, it lies well within the determinate region in the baseline specification. Figure 6.1(a) illustrates how unlikely is indeterminacy in the endogenous attention model; even for the Great Inflation period,

[31] See Christiano et al. (2005), for example.

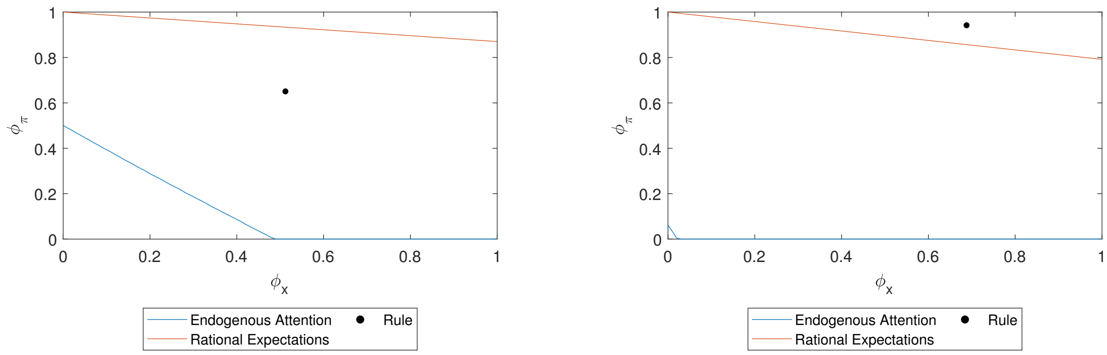
[32] Note that I hold ρ_i constant in either case and just vary ϕ_π and ϕ_x.

[33] Using the posterior mean as a point estimate minimises a quadratic loss function; see Herbst and Schorfheide (2015), p.41.

Table 3: Log Marginal Likelihood Comparison

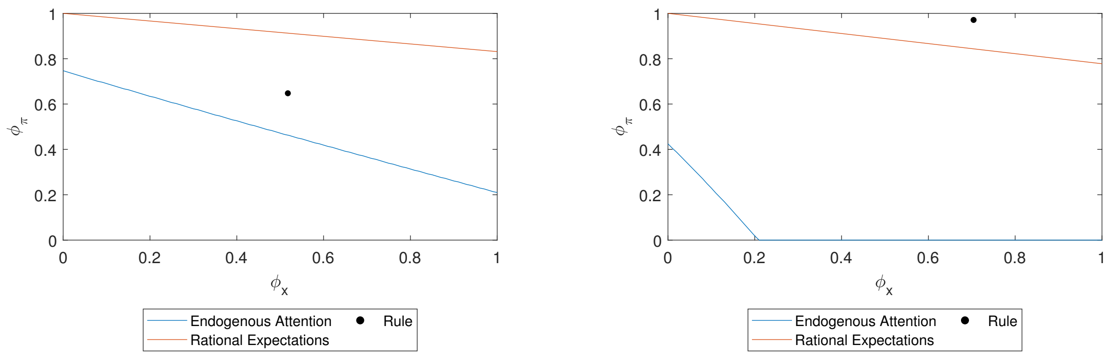
	Determinacy	Indeterminacy	Evidence
Baseline			
Great Inflation	-53.0	-67.8	Very Strong
Great Moderation (a)	84.4	23.6	Very Strong
Great Moderation (b)	64.0	20.2	Very Strong
Low Attn. Cost			
Great Inflation	-53.9	-57.2	Strong
Great Moderation (a)	82.4	43.7	Very Strong
Great Moderation (b)	62.3	35.9	Very Strong
Output Gap			
Great Inflation	-128.4	-141.7	Very Strong
Great Moderation (a)	15.6	-42.3	Very Strong
Great Moderation (b)	17.3	-25.4	Very Strong

"Evidence" refers to evidence against indeterminacy given by the Bayes factor, assessed using the scale given by Kass and Raftery (1995). Great Moderation (a) refers to the sample beginning in 1984:I, and (b) to the sample beginning 1990:I.



(a) Great Inflation (b) Great Moderation

Figure 6.1: Counterfactual Determinacy Regions: Baseline Specification



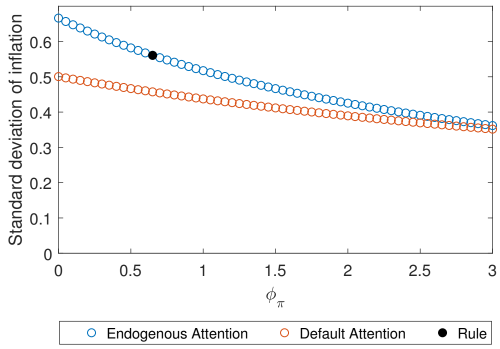
(a) Great Inflation (b) Great Moderation

Figure 6.2: Counterfactual Determinacy Regions: Low Attention Cost Specification

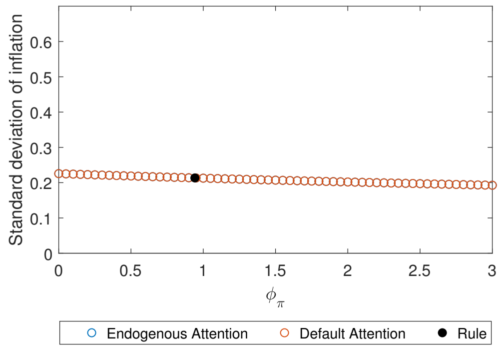
the indeterminacy region is small. Under the low attention cost, the indeterminacy region expands, but the estimated rule still lies well within the determinate region. For the post-1990 sub-sample, the point estimate for ϕ_π lies below unity across specifications. The point estimate for the rule still lies within the rational expectations determinacy region because of the high ϕ_x, but 39% of draws result in indeterminacy under rational expectations. However, under endogenous attention, the indeterminacy region is very small; volatility and hence attention were so low during this period that one would have to set ϕ_π and ϕ_x very close to 0 to allow indeterminacy.

Table 4: Posterior Distribution - Baseline Specification ($k_f = 1.5$)

	Great Inflation (1960:I to 1979:II)				Great Moderation (a) (1984:I to 2007:IV)				Great Moderation (b) (1990:I to 2007:IV)			
	Determinacy		Indeterminacy		Determinacy		Indeterminacy		Determinacy		Indeterminacy	
Parameter	Mean	90-pct. interval	Mean	90-pct. interval	Mean	90-pct. interval	Mean	90-pct. interval	Mean	90-pct. interval	Mean	90-pct. interval
ϕ_π	0.65	[0.55,0.76]	0.53	[0.52,0.62]	1.55	[1.00,2.17]	0.19	[0.08,0.32]	0.94	[0.49,1.47]	0.20	[0.08,0.33]
ϕ_x	0.51	[0.39,0.65]	0.48	[0.37,0.60]	0.73	[0.52,0.95]	0.64	[0.47,0.85]	0.69	[0.51,0.88]	0.49	[0.33,0.66]
θ	0.87	[0.83,0.92]	0.86	[0.82,0.89]	0.90	[0.87,0.93]	0.88	[0.86,0.90]	0.91	[0.88,0.93]	0.87	[0.79,0.90]
γ	2.90	[2.13,3.78]	3.42	[2.57,4.39]	4.02	[3.10,5.04]	4.91	[3.78,6.23]	3.49	[2.63,4.46]	4.22	[2.75,5.44]
ρ_η	0.74	[0.67,0.81]	0.86	[0.81,0.90]	0.66	[0.57,0.76]	0.86	[0.76,0.92]	0.63	[0.53,0.73]	0.84	[0.71,0.90]
ρ_z	0.80	[0.73,0.86]	0.76	[0.69,0.83]	0.91	[0.88,0.94]	0.88	[0.82,0.94]	0.90	[0.85,0.94]	0.86	[0.79,0.92]
ρ_i	0.56	[0.45,0.66]	0.56	[0.46,0.65]	0.84	[0.80,0.87]	0.84	[0.80,0.87]	0.83	[0.78,0.87]	0.81	[0.76,0.85]
σ_η^ε	0.12	[0.08,0.17]	0.08	[0.05,0.11]	0.08	[0.06,0.090]	0.04	[0.03,0.07]	0.08	[0.06,0.10]	0.06	[0.03,0.09]
σ_z^ε	0.48	[0.36,0.63]	0.62	[0.47,0.81]	0.28	[0.22,0.35]	0.39	[0.25,0.53]	0.28	[0.22,0.36]	0.38	[0.26,0.51]
σ_v^ε	0.16	[0.14,0.19]	0.16	[0.14,0.19]	0.12	[0.10,0.14]	0.12	[0.11,0.14]	0.10	[0.09,0.12]	0.10	[0.09,0.12]
σ_ζ			0.12	[0.06,0.20]			0.08	[0.05,0.13]			0.09	[0.05,0.15]
m_f	0.93	[0.85,0.98]	0.98	[0.96,0.99]	0.85	[0.85,0.89]	0.97	[0.95,0.98]	0.85	[0.85,0.88]	0.95	[0.85,0.98]
m_c	0.85	[0.85,0.85]	0.85	[0.85,0.85]	0.85	[0.85,0.85]	0.85	[0.85,0.85]	0.85	[0.85,0.85]	0.85	[0.85,0.85]
Log ML	-53.0		-67.8		84.4		23.6		64.0		20.2	

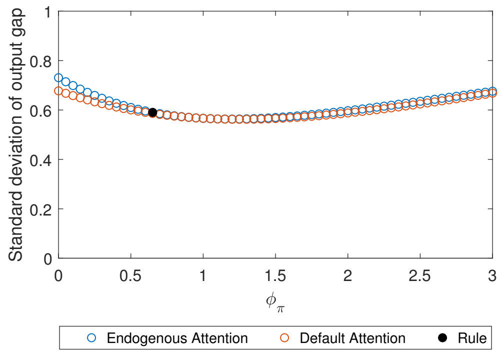


(a) Great Inflation

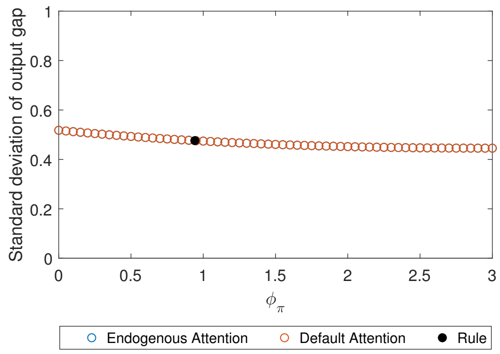


(b) Great Moderation

Figure 6.3: Counterfactual Inflation Volatility

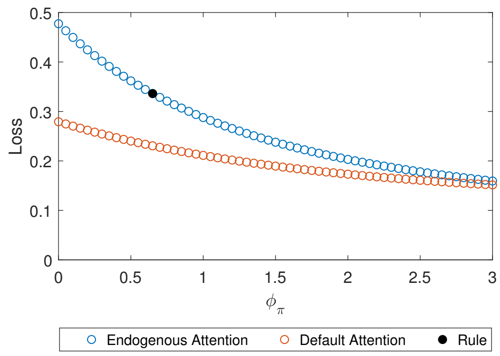


(a) Great Inflation

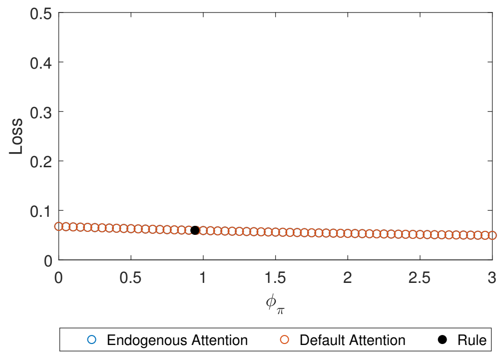


(b) Great Moderation

Figure 6.4: Counterfactual Output Gap Volatility



(a) Great Inflation



(b) Great Moderation

Figure 6.5: Counterfactual Central Bank Losses

I now conduct a counterfactual experiment by varying ϕ_π, holding the other parameters fixed. I show how the standard deviation of inflation and the output gap would have changed under different policy rules in figures 6.3 to 6.5. I also show central bank losses, which I calculate as:

$$L = V(\pi) + \xi V(x)$$

I use $\xi = 0.0625$.[34] In each figure, the black dot shows the volatility implied by the estimated rule. The blue series shows how volatility changes with ϕ_π under endogenous attention. The red series shows, for comparison, how volatility changes if attention is fixed at the default. For the Great Moderation attention is always at its default, so only the red series is visible. Table 6 compares volatility between the estimated ϕ_π for each period and $\phi_\pi = 2.0$.[35] To make a fair comparison across periods, as estimated policy rules differ, I also compare $\phi_\pi = 2.0$ and $\phi_\pi = 0.5$.

For the Great Inflation, inflation volatility and losses are high, and fall steeply as ϕ_π increases. For the Great Moderation, inflation volatility and losses are lower, and change proportionally less with ϕ_π. It makes sense that the absolute change in inflation volatility is greater for the Great Inflation, but why is it proportionally much larger? First, the estimated parameters imply a higher κ, allowing interest rate changes to have stronger effects on inflation. Second, ρ_η is higher during the Great Inflation, so raising ϕ_π tends to have a larger impact on inflation volatility. Consider the change in inflation volatility as ϕ_π moves from 2.0 to 0.5. In the Great Moderation, where attention is always at its default level, inflation volatility is just 8% higher when ϕ_π is set to 0.5. For the Great Inflation, even fixing attention at its default, the difference is 20%. However, a further effect results from shocks being large enough during the Great Inflation to push m_f above its default. Under endogenous attention, inflation volatility is 37% higher when ϕ_π is set to 0.5 than when it is set to 2.0. That is shown by the divergence of the red and the blue series in figure 6.3(a). As noted above, this divergence arises both because attention rises as ϕ_π decreases, and because when attention is higher the effects of ϕ_π on inflation volatility are more powerful. The results suggest these behavioural channels, which are only active when the economy is sufficiently volatile to push attention above the default, are quantitatively large.

Table 5: Counterfactual Macroeconomic Volatility and Central Bank Losses

	$\phi_\pi = 0.5$	Est. ϕ_π	$\phi_\pi = 2.0$	$\phi_\pi = 0.5$ vs. $\phi_\pi = 2.0$	Est. ϕ_π vs. $\phi_\pi = 2.0$
Standard Deviation π					
Great Moderation	0.219	0.213	0.202	8%	6%
Great Inflation (default)	0.467	0.457	0.389	20%	17%
Great Inflation	0.582	0.561	0.425	37%	32%
Standard Deviation x					
Great Moderation	0.493	0.476	0.452	9%	5%
Great Inflation (default)	0.601	0.586	0.588	2%	0%
Great Inflation	0.610	0.590	0.597	2%	-1%
Central Bank Loss					
Great Moderation	0.063	0.060	0.054	18%	12%
Great Inflation (default)	0.240	0.231	0.173	38%	33%
Great Inflation	0.362	0.336	0.203	78%	66%

The post-1990 sub-sample is used for the Great Moderation. The estimated ϕ_π is 0.65 for the Great Inflation, and 0.94 for the post-1990 period.

One might conclude from table 3 that there is no role for policy in the Great Inflation and Great Moderation. After all, the point estimate for ϕ_π changes little between the Great Inflation and the post-1990 sub-sample. The results in table 6 reject this view. Weak stabilization of inflation during the Great

[34] As noted in Debortoli et al. (2019), former Janet Yellen has suggested that the Fed places equal weight on annualized inflation and the unemployment gap. Here, π_t is quarter-on-quarter inflation, or one-quarter of annualized inflation. This translates into a weight of 1/16, or 0.0625, on the unemployment gap x_t relative to π_t.

[35] A response to inflation of 2.0 is roughly in line with the level estimated by Fernandez-Villaverde et al. (2010) for the majority of the 1980s.

Inflation resulted in losses that were 66% higher than they would have been if ϕ_π were set to 2.0. After 1990, the difference is just 12%. In absolute terms, differences are much larger. Raising ϕ_π to 2.0 would have reduced the standard deviation of quarterly inflation by 0.01 percentage point during the 1990s. For the Great Inflation, the reduction would have been 0.14 percentage points, or 0.54 percentage points in annualised terms. The policy rule implemented during the Great Inflation was a poor choice for that time; this was precisely the moment when a strong response to inflation would have been most beneficial. After 1990, the response to inflation may have been nearly as weak as during the Great Inflation. However, this mattered less for inflation volatility. Note that this discussion does not relate to optimal policy; it is about how large are the effects of moving the policy rule closer to the optimum.

6.2 Phillips and IS Curve Slopes

As a second application, I show that if households and firms discount the future, but one assumes that they are rational, then one will underestimate the slopes of the Phillips and IS curves. In the rational expectations model, the expectation terms generate amplification of persistence shocks. For example, the expectation term in the Phillips curve amplifies the effect of persistent changes in the output gap on inflation. If these expectations are dampened due to cognitive discounting, then the effect of the output gap on inflation is attenuated. As such, an econmetrician assuming rational expectations would mistakenly estimate a flatter Phillips curve.

Periods of macroeconomic stability should, according to the endogenous cognitive discounting model, lead households and firms to discount the future more heavily. This leads to the appearance of flatter Phillips and IS curves, if the econometrician is using a rational expectations model. This effect could have contributed to the apparent flattening of the Phillips curve often discussed in the literature.[36] This finding contributes to the growing literature that suggests that changes in expectation formation could account for the apparent flatness of the Phillips curve during the Great Moderation. See, for example, Lansing and Jorgensen (2022), who find a similar result holds using a signal-extraction model.

To illustrate this point, I consider three model specifications. The first has a default attention level of 0.85, as in the estimation section, and large enough shocks to push the firm attention level well above the default. The second specification has the same attention level, but smaller shocks, so that both firm and consumer attention are at the default. The third specification considers a lower default level of attention, of 0.60, and again uses smaller shocks so that attention remains at the default for both firms and consumers. For each specification, I simulate 100 datasets of 120 periods. I then estimate an endogenous cognitive discounting model and a rational expectations model on the resulting data using maximum likelihood. For comparison, I also estimate a model that assumes exogenous cognitive discounting. I then take the range of parameter estimates obtained, and take the (winsorized) mean and standard deviation.[37]

The results are shown in tables 6, 7, and 8. The rational expectations model gives a fairly accurate picture of the Phillips curve slope in the high shock specification. This is unsurprising, since the true value of m_f is close to unity, and so the rational expectations model is a fairly good approximation to the true model. In the lower shock specification, however, m_f is at its default, and so the rational expectations model is a poor approximation. In this case, the tendency is for the rational expectations model to underestimate the Phillips and IS curve slopes. This issue becomes even more severe in the model with a lower default level of attention, where the true values of m_f and m_c are just 0.6. In this case, the rational expectations model underestimate the IS and Phillips curve slopes by $40 - 50\%$.

The results also clearly illustrate the identification problem with the exogenous attention model. Note how large is the standard deviation of the point estimates for m_f and m_c. As these variables are bounded by the interval $[0, 1]$, the maximum standard deviation possible would be 0.5. As such, these parameters

[36] See McLeay and Tenreyro (2019) for an overview of the empirical evidence.

[37] I winsorize at the 5th and 95th percentiles for each of the parameters, to try to give an accurate representation of the typical results obtained.

Table 6: Distribution of Points Estimates Obtained Under MLE: High Default, Large Shocks

		Endogenous Attention		Exogenous Attention		Rational Expectations	
Parameter	True Value	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
ϕ_π	1.50	1.52	0.12	1.52	0.12	1.52	0.12
ϕ_x	0.25	0.24	0.08	0.24	0.08	0.24	0.07
θ	0.80	0.79	0.05	0.75	0.08	0.80	0.04
γ	2.00	2.36	1.20	1.66	0.71	2.83	1.50
ρ_η	0.80	0.79	0.05	0.78	0.06	0.79	0.05
ρ_z	0.90	0.88	0.06	0.88	0.07	0.88	0.04
ρ_i	0.80	0.80	0.02	0.80	0.02	0.80	0.02
σ_η^ε	0.20	0.22	0.05	0.33	0.16	0.18	0.04
σ_z^ε	0.30	0.33	0.06	0.49	0.25	0.24	0.05
σ_v^ε	0.20	0.20	0.01	0.20	0.01	0.20	0.01
m_f	0.95	0.93	0.05	0.65	0.41	1.00	0.00
m_c	0.85	0.85	0.00	0.59	0.40	1.00	0.00
κ	0.16	0.19	0.06	0.27	0.17	0.18	0.05
σ	0.50	0.50	0.19	0.70	0.27	0.43	0.18

Table 7: Distribution of Points Estimates Obtained Under MLE: high Default, Small Shocks

		Endogenous Attention		Exogenous Attention		Rational Expectations	
Parameter	True Value	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
ϕ_π	1.50	1.52	0.15	1.53	0.15	1.53	0.15
ϕ_x	0.25	0.24	0.07	0.24	0.07	0.25	0.07
θ	0.80	0.80	0.04	0.76	0.08	0.84	0.04
γ	2.00	2.57	1.64	1.74	0.86	3.25	1.89
ρ_η	0.80	0.78	0.05	0.78	0.06	0.79	0.05
ρ_z	0.90	0.87	0.06	0.88	0.07	0.87	0.04
ρ_i	0.80	0.80	0.02	0.80	0.02	0.80	0.02
σ_η^ε	0.10	0.10	0.02	0.15	0.07	0.07	0.01
σ_z^ε	0.15	0.16	0.03	0.25	0.13	0.12	0.03
σ_v^ε	0.10	0.10	0.01	0.10	0.01	0.10	0.01
m_f	0.85	0.87	0.04	0.53	0.45	1.00	0.00
m_c	0.85	0.85	0.00	0.58	0.41	1.00	0.00
κ	0.16	0.16	0.04	0.24	0.15	0.12	0.04
σ	0.50	0.50	0.21	0.71	0.32	0.39	0.18

could hardly be less precisely estimated. The imprecise estimates of the discounting parameters spills over into inaccurate estimates of the Phillips and IS curve slopes; note how much larger are the standard deviation of the point estimates of κ and σ under exogenous than endogenous attention. The problem of identifying the Phillips and IS curve slopes consequently cannot be easily resolved by simply adding exogenous discounting parameters into the model.

7 External Validation

The paucity of empirical estimates of the costs of attention in the microdata presents a challenge for estimating the endogenous attention model. My analysis shows that endogenizing the attention level is important empirically, and so calls for further empirical work to pin down attention costs. Until there is greater confidence in estimates of attention costs in the microdata, researchers may desire an alternative method for externally validating their results when estimating the endogenous attention model. This section demonstrates such a method, using the Great Inflation and Great Moderation estimation exercise as an example. Specifically, I verify whether the dynamics of firm cognitive discounting that drive my results match empirical data on expectation formation. Coibion and Gorodnichenko (2015a) study Survey of Professional Forecasters (SPF) data. They run regressions of the form:

$$x_{t+h} - F_t x_{t+h} = c + \beta(F_t x_{t+h} - F_{t-1} x_{t+h}) + \varepsilon_{t,h}$$

Table 8: Distribution of Points Estimates Obtained Under MLE: Low Default, Small Shocks

		Endogenous Attention		Exogenous Attention		Rational Expectations	
Parameter	True Value	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
ϕ_π	1.50	1.52	0.20	1.53	0.21	1.54	0.23
ϕ_x	0.25	0.25	0.10	0.25	0.10	0.25	0.10
θ	0.80	0.80	0.03	0.79	0.07	0.87	0.04
γ	2.00	2.18	0.77	2.19	1.15	4.20	1.92
ρ_η	0.80	0.78	0.05	0.78	0.05	0.80	0.06
ρ_z	0.90	0.88	0.05	0.88	0.06	0.88	0.04
ρ_i	0.80	0.80	0.03	0.79	0.03	0.79	0.02
σ^ε_η	0.10	0.10	0.01	0.11	0.05	0.05	0.01
σ^ε_z	0.15	0.16	0.01	0.17	0.09	0.07	0.02
σ^ε_v	0.10	0.10	0.01	0.10	0.01	0.10	0.01
m_f	0.60	0.60	0.00	0.46	0.46	1.00	0.00
m_c	0.60	0.60	0.00	0.55	0.42	1.00	0.00
κ	0.16	0.17	0.03	0.21	0.14	0.10	0.04
σ	0.50	0.51	0.16	0.57	0.26	0.28	0.12

Here, x_t is some macroeconomic variable, and $F_t x_{t+h}$ denotes the time t forecast at the h quarter horizon. This regression asks whether the forecast revision at time t predicts the ex-post error of the time t forecast. Under rational expectations, there is no relationship between forecast revisions and ex-post errors, and hence β is zero. An estimated β significantly different from zero then constitutes evidence against rational expectations. Coibion and Gorodnichenko refer to β as a measure of "informational rigidity".

Gabaix shows that under cognitive discounting, β should be different from zero. Specifically, at horizon h, β is bounded below by:

$$\underline{\beta}(m,h) = \frac{1-m^h}{m^h}$$

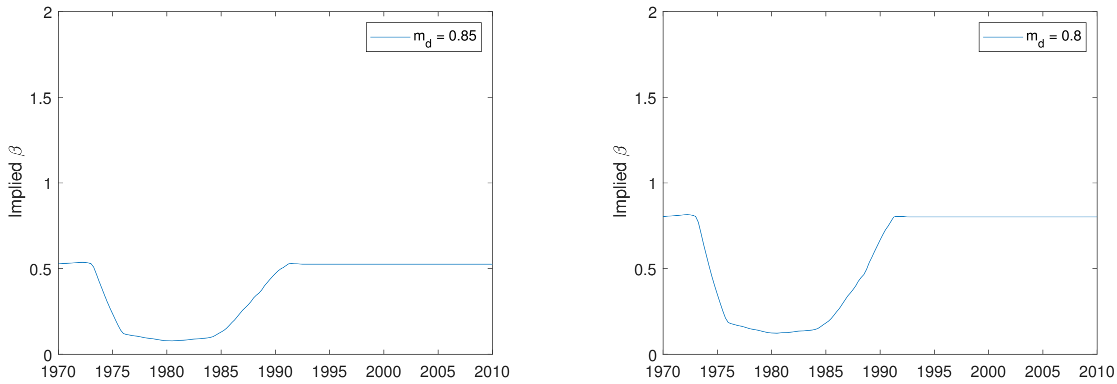
As such, the endogenous cognitive discounting model predicts that in periods of higher macroeconomic volatility, when m is higher, β should be lower.

Coibion and Gorodnichenko do indeed find evidence of this. Pooling forecast revisions and errors across a number of variables in the SPF, and across forecast horizons, they estimate β for each quarter from the beginning of the SPF sample in 1968, and take a smoothed average, shown in their figure 3. For comparison, they show the rolling 5 year standard deviation of US real GDP growth as a proxy for macroeconomic volatility. Informational rigidities are macroeconomic volatility are clearly inversely related; in the late 1970s, when macroeconomic volatility was high, informational rigidities declined. The estimated β then rose again as macroeconomic volatility declined in the late 1980s. While the professional forecasts in the SPF are not entirely comparable to firm forecasts, they should provide a guide to the direction and magnitude of changes in informational rigidities at least for large firms.

To compare their results with the predictions of my model, I estimate the model on rolling 40 quarter periods of data, and find the implied m_f at the posterior mode. I then calculate the implied lower bound on β. I note that Coibion and Gorodnichenko pool across horizons up to four periods ahead. To make things comparable, since $\underline{\beta}$ depends on h, I calculate an approximate lower bound for β in a pooled approach using an average of the lower bound for one to four quarter horizons:

$$\begin{aligned}\underline{\beta}_{pooled}(m) &= \sum_{h=1}^{4} \underline{\beta}(m,h) \\ &= \sum_{h=1}^{4} \frac{1-m^h}{m^h}\end{aligned}$$

Coibion and Gorodnichenko smooth their results, and I do the same by averaging across periods.



(a) $m_{f,d} = 0.85$ (b) $m_{f,d} = 0.8$

Figure 7.1: Implied Lower Bound on β

The results are shown in figure 7.1, using two different values of m_f. 7.1a uses $m_{f,d} = 0.85$, as in the baseline sepecification, and 7.1b uses a slightly lower value of 0.8. The degree of informational rigidity predicted by the endogenous cognitive discounting model decreases during the late 1970s, and subsequently increases after the mid-1980s, much as in the Coibion and Gorodnichenko findings. In the model, the lower bound on m_f is given by $m_{f,d}$, which is why the implied $\underline{\beta}_{pooled}$ reaches a maximum in the late 1980s. Nonetheless, the magnitude of the changes in $\underline{\beta}_{pooled}$ are in line with Coibion and Gorodnichenko's estimates of β, particularly for $m_{f,d} = 0.8$. As such, empirical data on expectations provides external validation for the direction and magnitudes of the changes in firms' cognitive discounting over this period, which are the key driver of my results.

8 Conclusions

Gabaix (2020) demonstrates that cognitive discounting has profound implications for macroeconomic theory, specifically in the new Keynesian framework. In this paper, I highlighted a number of cases in which theoretical and empirical analysis changes dramatically when the degree of discounting is endogenous.

I showed three cases in which endogenising attention makes a very considerable difference to the model's theoretical predictions. First, I showed that endogenizing the attention parameter qualitatively affects the determinacy condition. Under endogenous discounting, an indeterminate equilibrium always exists whenever the rational expectations Taylor principle is violated. However, my analysis shows that indeterminacy is path-dependent. If a determinate equilibrium exists, then it will often not be possible to reach the indeterminate equilibrium if the economy starts in a determinate regime. Whether an indeterminate equilibrium is possible in practice depends on the shocks hitting the economy. Second, under endogenous attention, the proportional effects of changes in the Taylor rule coefficients on inflation and output gap volatility depend on the size of the shocks hitting the economy. A stronger policy reaction to inflation has larger effects on inflation volatility when the shocks hitting the economy are larger. This suggests a stronger imperative for central banks to react aggressively to inflation when shocks are large. This result does not obtain under exogenous discounting or rational expectations. Third, exogenous discounting may understate the benefits of average inflation targeting. While AIT may not have large effects conditional on agents discounting the future a great deal, it may be necessary in order to induce agents to pay limited attention.

I then showed that endogenising the discount factor resolves a weak identification problem present in the exogenous discounting version of the model. I illustrated the utility of this in two empirical applications. First, one can correctly assess whether the economy was in a determinate or indeterminate regime. This led to a novel conclusion; there is strong evidence against indeterminacy as a cause of the Great Inflation. Second, one can conduct counterfactuals that are robust to the Lucas critique; one can assess the effects of counterfactual changes in the policy rule taking into account the effects on the equilibrium level of attention. Consistent with my theoretical analysis, I found that a stronger monetary policy reaction to inflation would have been highly beneficial during the Great Inflation period, but the strength of the policy reaction during the Great Moderation was less important in shaping the level of macroeconomic volatility. Third, one can correctly recover the slopes of the IS and Phillips curves. Endogenous discounting implies that econometricians assuming rational expectations may systematically underestimates the slopes of the curves, particularly during times of macroeconomic stability when the equilibrium level of attention is low. The slopes of the curves cannot be precisely recovered using the exogenous attention model, because of weak identification.

The framework used in this paper opens a broader research agenda. This paper focused on a three-equation new Keynesian setting. Incorporating capital investment, sticky wages, consumption habits or inflation indexation would presumably affect agents' incentives to pay attention and how changes in attention affect macroeconomic dynamics. Understanding the effects of endogenous attention in medium-scale models is consequently an important objective for future research.

References

Adam, K. (2009). Monetary Policy and Aggregate Volatility. *Journal of Monetary Economics*, **56**(S), 1-18.

Andrade, J., Cordeiro, P., and Lambais, G. (2019). Estimating a Behavioral New Keynesian Model. *Working Paper.*

Andre, P., Haaland, I., Roth, C., and Wohlfart, J. (2021). Inflation Narratives. *Working Paper.*

Andrews, I. (2018)."Valid Two-Step Identification-Robust Confidence Sets for GMM. *Review of Economics and Statistics.* **100**(2), 337-348.

Andrews, I., and Mikusheva, A. (2014), Weak Identification in Maximum Likelihood: A Question of Information. *American Economic Review*, **104**(5), 195-99.

Andrews, I., and Mikusheva, A. (2015). Maximum Likelihood Inference in Weakly Identified Dynamic Stochastic General Equilibrium Models. *Quantitative Economics*, **6**(1), 123-152.

Ball, L. and Mazumder, S. (2015). A Phillips Curve with Anchored Expectations and Short-Term Unemployment. *Working Paper.*

Benati, L. (2008). Investigating Inflation Persistence Across Monetary Regimes. *The Quarterly Journal of Economics*, **123**(3), 1005-1060.

Bernanke, B., Kiley, M., and Roberts, J. (2019). Monetary Policy Strategies for a Low-Rate Environment. *Working Paper.*

Bilbiie, F. (2021). Monetary Policy and Heterogeneity: An Analytical Framework. *Working Paper.*

Boivin, J., and Giannoni, M. (2006). Has Monetary Policy Become More Effective? *Review of Economics and Statistics*, **88**(3), 445-462.

Bullard, J., and Mitra, K. (2002). Learning about Monetary Policy Rules. *Journal of Monetary Economics*, **49**(6), 1105-1129.

Bundiato, F., Nakata, T., and Schmidt, S. (2021). Average Inflation Targeting and the Interest Rate Lower Bound. *Working Paper.*

Canova, F. and Ferroni, F. (2011). Multiple Filtering Devices for the Estimation of Cyclical DSGE Models. *Quantitative Economics*, **2**(1), 73-98.

Chau, V. (2020). Behavioral-Attention Phillips Curve: Theory and Evidences from Inflation Surveys. *Working Paper.*

Clarida, R., Gali, J., and Gertler, M. (2000). Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory. *Quarterly Journal of Economics*, **115**(1), 147-180.

Christiano, L., Eichenbaum, M., and Evans, C. (2005). Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. *Journal of Political Economy*, **113**(1), 1-45.

Cochrane, J. (2016). Comments on 'A Behavioral New Keynesian Model' by Xavier Gabaix. *Working Paper.*

Coibion, O., and Gorodnichenko, Y. (2015a). Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts. *American Economic Review*, **105**(8), 2644-2678.

Coibion, O., and Gorodnichenko, Y. (2015b). Is the Phillips Curve Alive and Well after All? Inflation Expectations and the Missing Disinflation. *American Economic Journal: Macroeconomics*, **7**(1), 197-232.

Coibion, O., Gorodnichenko, Y., and Ropele, T. (2020). Inflation Expectations and Firm Decisions: New Causal Evidence. *Quarterly Journal of Economics*, **135**(1), 165-219.

Debortoli, D., Kinill, K., Lindé, J., and Nunes, R. (2019). Designing a Simple Loss Function for Central Banks: Does a Dual Mandate Make Sense? *Economic Journal*, **129**(621), 2010-2038.

Del Negro, M., Giannoni, M., and Patterson, C. (2015). The Forward Guidance Puzzle. *Federal Reserve Bank of New York Staff Reports.*

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

Farhi, E., and Werning, I. (2019). Monetary Policy, Bounded Rationality, and Incomplete Markets. *American Economic Review*, **109**(11), 3887-3928.

Fernandez-Villaverde, J., Guerron-Quintana, P., and Rubio-Ramirez, J. (2010). Fortune or Virtue: Time-Variant Volatilities Versus Parameter Drifting in U.S. Data. *Working Paper.*

Fuhrer, J. (2006). Intrinsic and Inherited Inflation Persistence. *International Journal of Central Banking*, **2**(3).

Fuhrer, J., and Rudebusch, G. (2004). Estimating the Euler Equation for Output. *Journal of Monetary Economics*, **51**(6), 1133-1153.

Gabaix, X. (2014). A Sparsity-Based Model of Bounded Rationality. *The Quarterly Journal of Economics*, **129**(4), 1661-1710.

Gabaix, X. (2017). Behavioral Macroeconomics Via Sparse Dynamic Programming. *Working Paper.*

Gabaix, X. (2020). A Behavioral New Keynesian Model. *American Economic Review*, **110**(8), 2271-2327.

Gali, J. and Gertler, M. (1999). Inflation Dynamics: A Structural Econometric Analysis. *Journal of Monetary Economics*, **44**(2), 195-222.

Ganong, P., and Noel, P. (2017). Consumer Spending during Unemployment: Positive and Normative Implications. *Working Paper.*

Geweke, J. (1999). Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication. *Econometric Reviews*, **18**(1), 1-73.

Herbst, E., and Schorfheide, F. (2015). *Bayesian Estimation of DSGE Models.* Princeton University Press.

Hommes, C. (2018). Behavioral & Experimental Macroeconomics and Policy Analysis: A Complex Systems Approach. *Working Paper.*

Ilabaca, F., Meggiorini, G., and Milani, F. (2020). Bounded Rationality, Monetary Policy, and Macroeconomic Stability. *Economics Letters*, **186**.

International Monetary Fund (2013). The Dog that Didn't Bark: Has Inflation Been Muzzled or Was it Just Sleeping? *Working Paper.*

Iskrev, N. (2010a). Local Identification in DSGE Models. *Journal of Monetary Economics*, **57**(2), 189-202.

Iskrev, N. (2010b). Evaluating the Strength of Identification in DSGE Models: An A Priori Approach. *Working Paper.*

Justiniano, A., and Primiceri, G. (2008). The Time-Varying Volatility of Macroeconomic Fluctuations. *American Economic Review*, **98**(3), 603-641.

Kass, R., and Raftery, A. (1995). Bayes Factors. *Journal of the American Statistical Association*, **90**(430), 773-795.

Kiley, M., and Roberts, J. (2017). Monetary Policy in a Low Interest Rate World. *Brookings Papers on*

Economic Activity, **48**(1), 317-396.

Koop, G., Hashem Pesaran, M., and Smith, R. (2013). On Identification of Bayesian DSGE Models. *Journal of Business & Economic Statistics*, **31**(3), 300-314.

Lange, M., Zuhlke, D., Holz, O. and Villmann, T. (2014). Applications of l_p-Norms and their Smooth Approximations for Gradient Based Learning Vector Quantization. *European Symposium on Artificial Neural Networks Proceedings.*

Lansing, P., and Jorgensen, K. (2022). Anchored Inflation Expectations and the Slope of the Phillips Curve. *Working Paper.*

Lubik, T., and Schorfheide, F. (2003). Computing Sunspot Equilibria in Linear Rational Expectations Models. *Journal of Economic Dynamics and Control*, **28**(2), 273-285.

Lubik, T., and Schorfheide, F. (2004). Testing for Indeterminacy: An Application to U.S. Monetary Policy. *American Economic Review*, **94**(1), 190-217.

Lubik, T., and Marzo, M. (2021). Fiscal Policy Perceptions in a Behavioral New Keynesian Model. *Annals of Economics and Finance*, **22**(2), 255-287.

Lucas, R. (1976). Econometric Policy Evaluation: A Critique. *Carnegie-Rochester Conference Series on Public Policy*, **1**(1), 19-46.

Macowiak, B., and Wiederholt, M. (2009). Optimal Sticky Prices under Rational Inattention. *American Economic Review*, **99**(3), 769-803.

Macowiak, B., and Wiederholt, M. (2015). Business Cycle Dynamics under Rational Inattention. *Review of Economic Studies*, **82**(4), 1502-1532.

Mankiw, N., and Reis, R. (2002). Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve. *The Quarterly Journal of Economics*, **117**(4), 1295-1328.

Mankiw, N., and Reis, R. (2006). Sticky Information in General Equilibrium. *Working Paper.*

Meggiorini, G. (2021). Behavioural New Keynesian Models: An Empirical Assessment. *Working Paper.*

Meggiorini, G., and Milani, F. (2021). Behavioural New Keynesian Models: Learning vs. Cognitive Discounting. *Working Paper.*

McKay, A., Nakamura, E., and Steinsson, J. (2016). The Power of Forward Guidance Revisited. *American Economic Review*, **106**(10), 3133-58.

McLeay, M., and Tenreyro, S. (2019). Optimal Inflation and the Identification of the Phillips Curve. *Working Paper.*

Menkulasi, J. (2009). Rational inattention and changes in macroeconomic volatility. *Working Paper.*

Miranda, M. and Fackler, P. (2002). *Applied Computational Economics and Finance.* MIT Press.

Moberly, J. (2020). Rational Inattention and the Great Moderation. *Unpublished MPhil thesis.*

Nakamura, E. and Steinsson, J. (2008). Five Facts about Prices: A Reevaluation of Menu Cost Models. *Quarterly Journal of Economics*, **123**(4), 1415-1464.

Reis, R. (2006a). Inattentive Consumers. *Journal of Monetary Economics*, **53**(8), 1761-1800.

Reis, R. (2006b). Inattentive Producers. *The Review of Economic Studies*, **73**(3), 793-821.

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

Smets, F., and Wouters, R. (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review*, **93**(3), 586-606.

St Amant, P., and van Norden, S. (1997). Measurement of the Output Gap: A Discussion of Recent Research at the Bank of Canada. Technical Report of the Bank of Canada, **79**.

Stock, J., and Watson, M. (2002). Has the Business Cycle Changed and Why? *Working Paper*.

Woodford, M. (2001). The Taylor Rule and Optimal Monetary Policy. *American Economic Review*, **91**(2), 232-237.

Woodford, M. (2013). Macroeconomic Analysis Without the Rational Expectations Hypothesis. *Working Paper*.

Appendix A

Exogenous Attention

I begin by solving the system under exogenous attention, and then use those dynamics to formulate the g mapping. To solve the system, one can simply follow Lubik and Schorfheide (2003, 2004). This allows one to solve the system under both determinacy and indeterminacy. Denote by $\boldsymbol{s_t}$ a vector of endogenous variables, augmented by the rational expectations terms $E_t\pi_{t+1}$ and $E_t x_{t+1}$. Then the system can be written as:

$$\boldsymbol{\Gamma_0(M,\chi)s_t = \Gamma_1(\chi)s_{t-1} + \Psi\varepsilon_t + \Pi\eta_t}$$

Here, $\boldsymbol{\eta_t}$ is a vector of expectation errors, and $\boldsymbol{\varepsilon_t}$ is a vector comprising the innovations in the fundamental shock processes. Following Lubik and Schorfheide, one can write the solution to the model as:

$$\boldsymbol{s_t = Z_1(M,\chi)s_{t-1} + Z_2(M,\chi,Q)\varepsilon_t + Z_3(M,\chi)}\zeta_t$$

Under determinacy, the vector $\boldsymbol{Z_3(M)}$ is restricted to be a vector of zeros; sunspot shocks are ruled out by assumption because they have greater than unit root persistence. Under indeterminacy, however, sunspot shocks may affect the endogenous variables. Under determinacy, the matrix $\boldsymbol{Z_2(M,\chi,Q)}$ is uniquely determined by $\boldsymbol{M}$ and $\boldsymbol{\chi}$. Under indeterminacy, the response of the endogenous variables to fundamental shocks may vary. Throughout, I use Lubik and Schorfheide's (2003, 2004) continuity solution as a baseline.[38] The vector $\boldsymbol{Q}$ parameterises how responses differ from this continuity solution. Thus, under the continuity solution it is just a vector of zeros, and one can just write $\boldsymbol{Z_2}$ as a function of $\boldsymbol{M}$ and $\boldsymbol{\chi}$.

The dynamics above imply a particular variance-covariance matrix of the endogenous variables, which I will denote by $\boldsymbol{\Sigma_s(M,\chi)}$.

Consumption with a Time-Varying Discount Factor

As noted in the main text, the consumer's problem changes slightly when demand shocks are considered to be discount factor shocks (which stem from within the consumer problem) rather than technology shocks (which stem from outside the firm problem). Here, I derive an expression for $\hat{c}_t$ in the presence of discount factor shocks, which will be needed to calculate consumer attention. I use the same strategy as Gabaix (2020), modified to include a discount factor shock. I consider the deterministic version of the problem. Taking income (and hence labour supply) as given, the consumer's budget constraint can be solved forwards as follows:

$$C_t + \frac{C_{t+1}}{R_t} + \frac{C_{t+2}}{R_t R_{t+1}} + ... = Y_t + \frac{Y_{t+1}}{R_t} + \frac{Y_{t+2}}{R_t R_{t+2}} + ...$$

Using the consumption Euler equation then gives:

$$C_t = \left(1 + \beta_t^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}-1} + ...\right)^{-1} \left(Y_t + \frac{Y_{t+1}}{R_t} + \frac{Y_{t+2}}{R_t R_{t+1}} + ...\right)$$

Linearising then yields the following relationship.

$$\widehat{c}_t = \sum_{\tau \geq 0} \beta^\tau \left((1-\beta)\widehat{y}_{t+\tau} - \frac{\beta^2}{\gamma}\widehat{r}_{t+\tau} - \frac{1}{\gamma} z_{t+\tau}\right)$$

Consumers' Cognitive Discount Factor

The calculation of consumer attention is now slightly more complex because under indeterminacy and with interest rate smoothing the expectations of the endogenous variables cannot just be written as

[38] There is a slight complication in that one has to calculate the "nearest" point at the edge of the the determinacy region to calculate the continutiy solution. To get to the edge of the determinacy region one could adjust m_f or m_c instead of ϕ_π, but to be consistent with Lubik and Schorfheide I use ϕ_π.

linear functions of the contemporaneous shocks. I instead propose the following method. First, define the selection vectors $\boldsymbol{k_y}$, $\boldsymbol{k_r}$, and $\boldsymbol{k_z}$ such that:

$$\begin{aligned}\boldsymbol{k_y s_t} &= \widehat{y}_t \\ \boldsymbol{k_r s_t} &= \widehat{r}_t \\ \boldsymbol{k_z s_t} &= z_t\end{aligned}$$

Note that absent technology shocks $x_t = \widehat{y}_t$. The above expression can then be rewritten as:

$$\begin{aligned}\partial c_t/\partial m_c &= \frac{1}{m_c} E_t \sum_{\tau \geq 0} \tau (\beta m_c)^\tau \left((1-\beta)\boldsymbol{k_y} - \frac{\beta^2}{\gamma}\boldsymbol{k_r} - \frac{1}{\gamma}\boldsymbol{k_z} \right) \boldsymbol{s_{t+\tau}} \\ &= \frac{1}{m_c} E_t \sum_{\tau \geq 0} \tau (\beta m_c)^\tau \left((1-\beta)\boldsymbol{k_y} - \frac{\beta^2}{\gamma}\boldsymbol{k_r} - \frac{1}{\gamma}\boldsymbol{k_z} \right) \boldsymbol{Z_1(M,\chi)^\tau s_t}\end{aligned}$$

For compactness of notation, define:

$$\boldsymbol{k_c} = (1-\beta)\boldsymbol{k_y} - \frac{\beta^2}{\gamma}\boldsymbol{k_r} - \frac{1}{\gamma}\boldsymbol{k_z}$$

Noting the presence of the τ in the summation, and recalling that the eigenvalues of $\boldsymbol{Z_1(M,\chi)}$ lie within the unit circle, one can write this as:

$$\partial c_t/\partial m_c = \boldsymbol{k_c}(\boldsymbol{I} - \beta m_c \boldsymbol{Z_1(M,\chi)})^{-2} \beta \boldsymbol{Z_1(M,\chi) s_t}$$

I take the expectation of the square, and evaluate at the default as explained above. I obtain the key quantity in the attention function:

$$E[(\partial c_t/\partial m_c)^2] = \beta^2 \boldsymbol{k_c}(\boldsymbol{I} - \beta m_{c,d}\boldsymbol{Z_1(M,\chi)})^{-2} \boldsymbol{Z_1(M,\chi)\Sigma_s(M,\chi)}\ \boldsymbol{Z_1(M,\chi)'} \left((\boldsymbol{I} - \beta m_{c,d}\boldsymbol{Z_1(M,\chi)})^{-2} \right)' \boldsymbol{k_c'}$$

Note that the dynamics of the model are taken as given, so that $\boldsymbol{Z_1(M,\chi)}$, and indeed $\boldsymbol{\Sigma_s(M,\chi)}$, are unaffected by evaluating at the default.

Firms' Cognitive Discount Factor

I now show the full derviation for equation the firms' choice of cognitive discount factor. I begin with the following equation from Gabaix (2020), which gives the optimal choice of relative price q for firms resetting prices in time t:[39]

$$\begin{aligned}q_t &= (1-\beta\theta) \sum_{\tau \geq 0} (\beta\theta m_f)^\tau E_t(\pi_{t+1} + ... + \pi_{t+\tau} + \widehat{mc}_{t+\tau}) \\ \partial q_t/\partial m_f &= (1-\beta\theta)\frac{1}{m_f} \sum_{\tau \geq 0} \tau (\beta\theta m_f)^\tau E_t(\pi_{t+1} + ... + \pi_{t+\tau} + \widehat{mc}_{t+\tau})\end{aligned}$$

Note that no adjustment to the firm's attention problem is required because: (i) linearisation means that fluctuations in the discount factor do not enter into the expression for the optimal relative price, and (ii) the cost-push shock η_t originates outside the firm's problem.

I define selection matrices $\boldsymbol{k_\pi}$ and $\boldsymbol{k_{mc}}$ such that:

$$\begin{aligned}\boldsymbol{k_\pi s_t} &= \pi_t \\ \boldsymbol{k_{mc} s_t} &= \widehat{mc}_t\end{aligned}$$

[39] See Gabaix (2020) equation (173).

I then follow the same procedure as before. To lighten the notation, I denote $\boldsymbol{Z_1(m)}$ simply as $\boldsymbol{Z_1}$, though of course this still depends on the choices of attention.

$$\begin{aligned}
\partial q_t/\partial m_f &= (1-\beta\theta)\frac{1}{m_f}\sum_{\tau\geq 0}\tau(\beta\theta m_f)^{\tau}E_t(\boldsymbol{k_\pi s_{t+1}}+...+\boldsymbol{k_\pi s_{t+\tau}}+\boldsymbol{k_{mc}s_{t+\tau}})\\
&= (1-\beta\theta)\frac{1}{m_f}\sum_{\tau\geq 0}\tau(\beta\theta m_f)^{\tau}(\boldsymbol{k_\pi Z_1}(\boldsymbol{I}+...+\boldsymbol{Z_1^{\tau-1}})+\boldsymbol{k_{mc}Z_1})\boldsymbol{s_t}\\
&= (1-\beta\theta)\frac{1}{m_f}\sum_{\tau\geq 0}\tau(\beta\theta m_f)^{\tau}(\boldsymbol{k_\pi Z_1}(\boldsymbol{I}-\boldsymbol{Z_1})^{-1}(\boldsymbol{I}-\boldsymbol{Z_1}^{\tau})+\boldsymbol{k_{mc}Z_1})\boldsymbol{s_t}\\
&= (1-\beta\theta)\frac{1}{m_f}\sum_{\tau\geq 0}\tau(\beta\theta m_f)^{\tau}(\boldsymbol{k_\pi Z_1}(\boldsymbol{I}-\boldsymbol{Z_1})^{-1}-\boldsymbol{k_\pi Z_1}(\boldsymbol{I}-\boldsymbol{Z_1})^{-1}\boldsymbol{Z_1}^{\tau})+\boldsymbol{k_{mc}Z_1})\boldsymbol{s_t}
\end{aligned}$$

One then arrives at the following:

$$\begin{aligned}
\partial q_t/\partial m_f &= (1-\beta\theta)\beta\theta\boldsymbol{P s_t}\\
E[(\partial q_t/\partial m_f)^2] &= (1-\beta\theta)^2(\beta\theta)^2\boldsymbol{P\Sigma_s P'}
\end{aligned}$$

The matrix $\boldsymbol{P}$ is defined as follows (noting that the termabove is evaluated at $m_{f,d}$):

$$\begin{aligned}
\boldsymbol{P} := \boldsymbol{k_\pi Z_1(I-Z_1)^{-1}}((\boldsymbol{I}-\beta\theta m_{f,d}\boldsymbol{I})^{-2} &- (\boldsymbol{I}-\beta\theta m_{f,d}\boldsymbol{Z_1})^{-2}\boldsymbol{Z_1})\\
&+\boldsymbol{k_{mc}}(\boldsymbol{I}-\beta\theta m_{f,d}\boldsymbol{Z_1})^{-2}\boldsymbol{Z_1}
\end{aligned}$$

Regime Switching

State N

In the regime switching model, I focus on determinate equilibria. This allows me to adopt a somewhat simpler method of solving the model. To solve the model with exogenous attention, one can iterate to find the minimum state variable solution. Taking as given the level of attention, the state N model can be written as:

$$\boldsymbol{A_N s_t} = \boldsymbol{B_N s_{t-1}} + \boldsymbol{C_N}E_t\boldsymbol{s_{t+1}} + \boldsymbol{D_N\varepsilon_t} + \boldsymbol{E_N}$$

Here, note that $\boldsymbol{s_t}$ does not contain expectation terms. In state N, the vector $\boldsymbol{E_N}$ is just a vector of zeros, giving a steady state that is just a vector of zeros. One can use a straightforward iteration to solve for the minimum state variable solution to the model:

$$\boldsymbol{s_t} = \boldsymbol{F_N s_t} + \boldsymbol{G_N\varepsilon_t}$$

With the solution in hand, one can then straightforwardly obtain the covariance matrix $\boldsymbol{\Sigma_{s.N}}$.

To calculate the implied level of attention, one can use the same method as above, with one slight difference. Here, $\boldsymbol{s_t}$ does not contain the expectation terms. For consumers, one has to calculate:

$$\hat{c}_t = -\sigma\boldsymbol{k_i}\sum_{h\geq 0}m_c^h E_t\boldsymbol{s_{t+h}} + \sigma\boldsymbol{k_p i}\sum_{h\geq 0}m_c^h E_t\boldsymbol{s_{t+h}} + \beta^{-2}\sigma\boldsymbol{k_z}\sum_{h\geq 0}m_c^h E_t\boldsymbol{s_{t+h}}$$

Calculating the expectation in state N is straightforward, because it is an absorbing state. One obtains:

$$\hat{c}_t = -\sigma(\boldsymbol{k_i}-\beta^{-2}\boldsymbol{k_z})(\boldsymbol{I}-m_c\boldsymbol{F_N})^{-1}\boldsymbol{s_t} + \sigma\boldsymbol{k_p i}(\boldsymbol{I}-m_c\boldsymbol{F_N})^{-1}\boldsymbol{F_N s_t}$$

One could differentiate this expression with respect to m_c analytically. I will actually proceed numerically, because the analytical solution becomes too complicated in state E. So, I write this expression as:

$$\hat{c}_t = \boldsymbol{\Omega_N}(m_c)\boldsymbol{s_t}$$

To calculate the expectation of the square of the derivative with respect to m_c, evaluated at the default:

$$\hat{c}_t = \left(\frac{\boldsymbol{\Omega_N}(m_{c,d}+\epsilon) - \boldsymbol{\Omega_N}(m_{c,d})}{\epsilon} \right) \boldsymbol{\Sigma_{s,N}} \left(\frac{\boldsymbol{\Omega_N}(m_{c,d}+\epsilon) - \boldsymbol{\Omega_N}(m_{c,d})}{\epsilon} \right)'$$

The calculation proceeds in the same fashion for firms. To calculate the equilibrium level of attention, I proceed numerically using a minimisation routine in MATLAB. Multiplicity is generally not a concern in state N, but if desired one can use a routine such as *fmincon* to search for equilibria in different intervals of m_f and m_c.

State E

In state E, matters are complicated by the fact that this is not an absorbing state. To solve the model for a given set of attention parameters, write the model as:

$$\boldsymbol{A_E s_t} = \boldsymbol{B_E s_{t-1}} + \boldsymbol{C_E} E_t \boldsymbol{s_{t+1}} + \boldsymbol{D_E \varepsilon_t} + \boldsymbol{E_E}$$

Then note that $E_t \boldsymbol{s_{t+1}}$ is given by:

$$\begin{aligned} E_t \boldsymbol{s_{t+1}} &= p E_t \boldsymbol{s^E_{t+1}} + (1-p) E_t \boldsymbol{s^N_{t+1}} \\ &= p E_t \boldsymbol{s^E_{t+1}} + (1-p) \boldsymbol{F_N s^E_t} \end{aligned}$$

Here, $E_t \boldsymbol{s^E_{t+1}}$ denotes the expectation condition on being in state E. Then one can rewrite the model as:

$$\boldsymbol{A_E s^E_t} = \boldsymbol{B_E s^E_{t-1}} + p \boldsymbol{C_E} E_t \boldsymbol{s^E_{t+1}} + (1-p) \boldsymbol{C_E F_N s^E_t} + \boldsymbol{D_E \varepsilon_t} + \boldsymbol{E_E}$$

One can then again proceed using an iteration to solve for:

$$\boldsymbol{s^E_t = F_E s^E_t + G_E \varepsilon_t + H_E}$$

I then use this equation to solve for the steady state $\boldsymbol{\bar{s}_E}$. One can then write:

$$\boldsymbol{s^E_t - \bar{s}_E = F_E (s^E_t - -\bar{s}_E) + G_E \varepsilon_t}$$

One can use this to solve for the variance-covariance matrix $\boldsymbol{\Sigma_{s,E}}$.

Solving for the optimal choice of attention is now much more complex, because now the expectation $\boldsymbol{s_{t+h}}$ depends on whether the economy switched into state N, and if so at what time. Formally:

$$E_t \boldsymbol{s_{t+h}} = p^h \boldsymbol{F^h_E (s_t - \bar{s}_E) + \bar{s}_E} + (1-p)(p^{h-1} \boldsymbol{F_N (F^{h-1}_E (s_t - \bar{s}_E) + \bar{s}_E)} + p^{h-2} \boldsymbol{F^2_N (F^{h-2}_E (s_t - \bar{s}_E) + \bar{s}_E)} + ... + p^0$$

Combining this with the expression for $\hat{c}_t$ given above, one obtains:

$$\hat{c}_t = \boldsymbol{\Omega_{1,E}}(m_c) \boldsymbol{\bar{s}_E} + \boldsymbol{\Omega_{2,E}}(m_c) \boldsymbol{(s_t - \bar{s}_E)}$$

Where:

$$\begin{aligned} \boldsymbol{\Omega_{1,E}}(m_c) = {} & -\sigma(\boldsymbol{k_i} - \beta^{-2} \boldsymbol{k_z}) \sum_{h} \geq 0 m_c^h \left(p^h \boldsymbol{F^h_E} + (1-p) \sum_{s=1}^{h} p^{h-s} \boldsymbol{F^s_N F^{h-s}_E} \right) + \\ & \sigma \boldsymbol{k_\pi} \sum_{h \geq 0} m_c^h \left(p^{h+1} \boldsymbol{F^{h+1}_E} + (1-p) \sum_{s=1}^{h+1} p^{h+1-s} \boldsymbol{F^s_N F^{h+1-s}_E} \right) \end{aligned}$$

And:

$$\begin{aligned} \boldsymbol{\Omega_{2,E}}(m_c) = {} & -\sigma(\boldsymbol{k_i} - \beta^{-2} \boldsymbol{k_z} - p \boldsymbol{k_p i F_E})(\boldsymbol{I} - m_c p \boldsymbol{F_E})^{-1} - \\ & \sigma(1-p)(\boldsymbol{k_i} - \beta^{-2} \boldsymbol{k_z}) \sum_{h \geq 0} m_c^h \sum_{s=1}^{h} p^{h-s} \boldsymbol{F^s_N F^{h-s}_E} + \sigma(1-p) \boldsymbol{k_\pi} \sum_{h \geq 0} m_c^h \sum_{s=1}^{h+1} p^{h+1-s} \boldsymbol{F^s_N F^{h+1-s}_E} \end{aligned}$$

These matrices can be calculated numerically, summing from $h = 0$ to $h = H$ for some very large H. One can then solve for the expectation of the square of the derivative, and hence the optimal choice of attention, as above. The derivation for firms proceeds in the same fashion. To find attention equilibria, I once again use a MATLAB minimisation routine. Here, multiplicity is much more of a concern. To deal with this, one can either: (i) visualize the set of equilibria, as I do in the main body of the paper, to check for multiplicity; or (ii) use a minimisation routine to check for equilibria in a set of specific intervals of m_f and m_c.

Appendix B

Consumer Attention

I now briefly discuss the determinants of consumer attention. In the model, consumer attention is largely driven by real interest rates rather than output. The coefficient on output in the selection matrix $\boldsymbol{k_c}$ is $(1-\beta)$, which is the marginal propensity to consume today out of a change in the discounted value of lifetime income. In an infinite horizon optimization problem, persistent deviations in real interest rates may cause large changes in consumption. Small changes in income, even if persistent, tend to have smaller effects.

For cost-push or technology shocks, the volatility of the real interest rate rises as ϕ_π increases. In the case of cost-push shocks, inflation stabilization is achieved by causing the real rate to deviate from its natural level, causing changes in the output gap and real wages that partially offset the exogenous change in marginal costs. For technology shocks, stabilisation of inflation is achieved by stabilising the real rate around its natural level. That involves causing larger deviations from steady state, which is what matters for the consumer's problem. For some combinations of shock variances, greater stabilization of inflation may therefore raise consumer attention, potentially increasing output gap volatility. For discount factor shocks, a higher ϕ_π increases real interest rate volatility by stabilising the real rate around a fluctuating natural level. In this case, however, that lowers consumer attention, because the shock comes from within the consumer's problem. The consumer cares about fluctuations in $\beta_t R_t$, and this term is less volatile if ϕ_π is higher. Under endogenous attention, then, the source of shocks matters for policy analysis. For policy shocks, greater stabilisation of inflation partially reverses the initial shock to the real interest rate, and so lowers consumer attention. The effect of ϕ_x is ambiguous. For cost-push, discount factor, or policy shocks, greater output gap stabilisation lowers consumer attention. For technology shocks, though, it increases attention; stabilisation occurs by pushing the real interest rate towards its natural level, and hence further from its steady state.

Effects of Attention on Macroeconomic Volatility

As I noted in the main text, how far changes in attention affect inflation volatility depends on the model's other parameters. To see this, note that under determinacy inflation can be written as: $\pi_t = \psi_{\pi\eta}\eta_t$. $\psi_{\pi\eta}$ can be solved for using the method of undetermined coefficients, to give:

$$\psi_{\pi\eta} = \frac{(1 - M_c\rho_\eta + \sigma\phi_x)}{(1 - M_c\rho_\eta + \sigma\phi_x)(1 - \beta M_f\rho_\eta) + \kappa\sigma(\phi_\pi - \rho_\eta)}$$

The derivative with respect to M_f is:

$$\partial\psi_{\pi\eta}/\partial M_f = \beta\rho_\eta\psi_{\pi\eta}^2$$

The effect of changes in M_f on inflation volatility consequently tend to be larger when shocks are more persistent, since an increase in ρ_η also increases $\psi_{\pi\eta}$.

Appendix C

Identification without Interest Rate Smoothing

For the model without interest rate smoothing, applying the Iskrev (2010) procedure yields the following results regarding point identification:

- If $\rho_\eta = \rho_z = \rho_v$, point identification fails even if M^f and M^c are known.
- If $\rho_\eta = \rho_z$, the model is identified if either M^f or M^c is known, but not if both are estimated.
- If $\rho_v = \rho_\eta$ or $\rho_v = \rho_z$ then point identification fails if either M^f or M^c has to be estimated.

If M^f and M^c are known, point identification only fails if all three autocorrelations are equal.

Appendix D

Firm Losses under Cognitive Discounting

Under Calvo pricing, the expected real-terms that firms make when they reset prices (until the next price reset) is:

$$D = \sum_{\tau \geq 0} \theta^\tau Q_{t,t+\tau} \left(\left(\frac{P_i}{P_{t+\tau}} \right)^{1-\epsilon} - \left(\frac{P_i}{P_{t+\tau}} \right)^{-\epsilon} MC^r_{t+\tau} \right)$$

Denote D^* the profits if P_i is set to the optimum P^*. Denote $\tilde{D}$ the profits under some alternative $\tilde{P}$. Now the expected loss in profits as a percentage of (optimum) sales is:

$$\begin{aligned} \frac{D^* - \tilde{D}}{S^*} &\approx \frac{\frac{\partial D}{\partial P_i}(\tilde{P} - P^*) + \frac{1}{2}\frac{\partial^2 D}{\partial P_i^2}(\tilde{P} - P^*)^2}{S^*} \\ &= \frac{\frac{1}{2}\frac{\partial^2 D}{\partial P_i^2}(\tilde{P} - P^*)^2}{S^*} \end{aligned}$$

Derivatives are evaluated at the optimum, which is why the first order term disappears. Expected sales are, in real terms:

$$S = \sum_{\tau \geq 0} \theta^\tau Q_{t,t+\tau} Y_{t+\tau} \left(\frac{P_i}{P_{t+\tau}} \right)^{1-\epsilon}$$

Next, I find the optimal price. The first-order condition is:

$$\frac{\partial D}{\partial P_i} = \sum_{\tau \geq 0} \theta^\tau Q_{t,t+\tau} Y_{t+\tau} \left((1-\epsilon) \frac{P_i^{-\epsilon}}{P_{t+\tau}^{1-\epsilon}} + \epsilon \frac{P_i^{-\epsilon-1}}{P_{t+\tau}^{-\epsilon}} MC^r_{t+\tau} \right)$$

The optimum price P^* equates this first-order condition to zero. Hence:

$$0 = \sum_{\tau \geq 0} \theta^\tau Q_{t,t+\tau} Y_{t+\tau} \left((1-\epsilon) \frac{(P^*)^{-\epsilon}}{P_{t+\tau}^{1-\epsilon}} + \epsilon \frac{(P^*)^{-\epsilon-1}}{P_{t+\tau}^{-\epsilon}} MC^r_{t+\tau} \right)$$

Rearranging gives:

$$(\epsilon - 1) \sum_{\tau \geq 0} \theta^\tau Q_{t,t+\tau} Y_{t+\tau} \left(\frac{(P^*)^{1-\epsilon}}{P_{t+\tau}^{1-\epsilon}} \right) = \epsilon \sum_{\tau \geq 0} \theta^\tau Q_{t,t+\tau} Y_{t+\tau} \left(\frac{(P^*)^{-\epsilon}}{P_{t+\tau}^{-\epsilon}} MC^r_{t+\tau} \right)$$

Denote these quantities as:

$$(\epsilon - 1)\zeta^* = \epsilon \omega^*$$

Now note that:

$$\begin{aligned}D^* &= \zeta^* - \omega^* \\ &= \frac{\epsilon}{\epsilon-1}\omega^* - \omega^* \\ &= \frac{1}{\epsilon-1}\omega^*\end{aligned}$$

Sales are:

$$S^* = \zeta^* = \frac{\epsilon}{\epsilon-1}\omega^*$$

Next, I find the second derivative:

$$\frac{\partial^2 D}{\partial P_i^2} = \sum_{\tau\geq 0}\theta^\tau Q_{t,t+\tau}Y_{t+\tau}\left(\epsilon(\epsilon-1)\frac{P_i^{-\epsilon-1}}{P_{t+\tau}^{1-\epsilon}} - \epsilon(\epsilon+1)\frac{P_i^{-\epsilon-2}}{P_{t+\tau}^{-\epsilon}}MC_{t+\tau}^r\right)$$

Evaluated at the optimum:

$$\frac{\partial^2 D}{\partial P_i^2} = \sum_{\tau\geq 0}\theta^\tau Q_{t,t+\tau}Y_{t+\tau}\left(\epsilon(\epsilon-1)\frac{(P^*)^{-\epsilon-1}}{P_{t+\tau}^{1-\epsilon}} - \epsilon(\epsilon+1)\frac{(P^*)^{-\epsilon-2}}{P_{t+\tau}^{-\epsilon}}MC_{t+\tau}^r\right)$$

Mutiply both sides by $(P^*)^2$:

$$\begin{aligned}(P^*)^2\frac{\partial^2 D}{\partial P_i^2} &= \sum_{\tau\geq 0}\theta^\tau Q_{t,t+\tau}Y_{t+\tau}\left(\epsilon(\epsilon-1)\frac{(P^*)^{1-\epsilon}}{P_{t+\tau}^{1-\epsilon}} - \epsilon(\epsilon+1)\frac{(P^*)^{-\epsilon}}{P_{t+\tau}^{-\epsilon}}MC_{t+\tau}^r\right) \\ &= \epsilon(\epsilon-1)\zeta^* - \epsilon(\epsilon+1)\omega^* \\ &- \epsilon^2\omega^* - \epsilon^2\omega^* - \epsilon e\omega^* \\ &= -\epsilon\omega^*\end{aligned}$$

As a consequence:

$$\begin{aligned}\frac{D^* - \tilde{D}}{S^*} &= \frac{1}{2}\frac{(P^*)^2\frac{\partial^2 D}{\partial P_i^2}}{S^*}\left(\frac{\tilde{P}-P^*}{P^*}\right)^2 \\ &= \frac{1}{2}\frac{-\epsilon\omega^*}{\frac{\epsilon}{\epsilon-1}\omega^*}\left(\frac{\tilde{P}-P^*}{P^*}\right)^2 \\ &= -\frac{1}{2}(\epsilon-1)\left(\frac{\tilde{P}-P^*}{P^*}\right)^2\end{aligned}$$

In expectation, the loss is then a constant times the expectation of the square of the percentage deviation in the pricing decision from the optimum. Suppose that the firm is not paying full attention to the future, and doing so changes the square of its pricing decision by $1.5\%^2$ on average. Then, using the above equation, and specifying $\epsilon = 10$, one obtains an expected loss of $c.0.1\%$ of sales.

Appendix E

Algorithm for Bayesian Estimation

My process for estimating the model broadly follows techniques outlined in Lubik and Schorfheide (2004) and Herbst and Schorfheide (2015), adapting these methods to allow for endogenous attention.

1. I write a fixed point iteration procedure to find the equilibrium level of attention for a given set of parameters. An explanation of this procedure is given in Appendix D.

2. I write a function which calculates the log-likelihood of a given set of parameters under determinacy or indeterminacy, using the Kalman filter. This log-likelihood takes into account the equilibrium level of attention implied by that set of parameters.

3. I combine the likelihood with the prior to compute the posterior.

4. I use a numerical procedure to find the mode of the posterior distribution within the determinate region and the indeterminate region. In order to force the function to remain within the determinate or indeterminate region, I use a penalty function.

5. I calculate the inverse Hessian of the posterior distribution at the mode. If the mode is close to the boundary of the region, and so affected by the penalty function, I find the Hessian at the mode under an alternative specification for attention, and/or an alternative sub-sample. I scale the Hessian by a constant c^2. I tried to choose c to give a c.30-40% acceptance rate.[40]

6. I use the mode and scaled inverse Hessian as inputs for a random walk Metropolis Hastings algorithm to find the unconditional posterior distribution of each parameter. I take 110,000 draws, discarding the first 10,000 as burn-in. I find the mean and 90 percent credibility interval. I use Geweke's (1999) modified harmonic mean to estimate the marginal likelihood.

Solving Numerically for Attention Equilibrium

As illustrated by the charts in section 4, each of $g^D(m_f, \boldsymbol{\chi})$ and $g^I(m_f, \boldsymbol{\chi})$ tend to have a unique fixed point (where such a fixed point exists at all). It is also worth noting that, if the fixed points of each of these functions are indeed unique, where they exist, then both of these functions have fixed points where the function crosses from above the 45 degree line to below it. A final feature worth noting is that the determinate and indeterminate regions for m_f are each convex sets, divided by a unique cut-off point m_f^* for which $h(m_f^*, \widehat{\boldsymbol{\chi}}) = 1$.

These features are helpful when finding fixed points numerically. I use a very simple fixed point iteration algorithm.[41] Assume some starting value for the m_f vector, denoted $m_{f,0}$, I perform the following iteration until convergence.

$$m_{f,n+1} = (1-c) \times m_{f,n} + c \times g(m_{f,n}, \widehat{\boldsymbol{\chi}})$$

Here, $c \in (0, 1]$ is an adjustment factor. To find a determinate equilibrium, I use $m_{f,0} = m_{f,d}$. It may be that this point is a fixed point. If there is no fixed point at $m_{f,d}$, then $g^D(m_{f,d}, \widehat{\boldsymbol{\chi}}) > m_{f,d}$, and so $m_{f,n}$ increases. Because a fixed point will involve the $g^D(m_{f,d}, \widehat{\boldsymbol{\chi}})$ crossing the 45 degree line from above, if $m_{f,n}$ is below the equilibrium level, it will increase, while if it is above the equilibrium level, it will decrease. As such, if such a fixed point does exist, I find that $m_{f,n}$ typically converges to it, provided c is sufficiently small. If there is no determinate equilibrium, then $m_{f,n}$ will eventually exceed m^{f*}, at which point the algorithm stops. For the indeterminate case, the same features of the $g^I(m^f, \widehat{\boldsymbol{\chi}})$ function help to find the fixed point. Starting at $m_{f,0} = 1$, the algorithm decreases $m_{f,n}$ until the fixed point is reached.

For the two-dimensional case, I use the same solution method as for the one-dimensional case, except that now the g function is bivariate rather than univariate. To assess attainability of the indeterminate equilibrium, I search for an equilibrium in the first period of indeterminacy. In equation (**??**), that means τ is set equal to 1. Strictly, this is a necessary condition for attainability. That said, as noted in the univariate examples in section 4, if g^I is non-increasing in the attention variables then it may be a necessary and sufficient condition.

[40] This acceptance rate is achieved in most sub-samples for most specifications, but the acceptance rate did turn out slightly below this range in some indeterminate sub-samples.

[41] A brief overview of fixed-point iteration is given in Miranda and Fackler (2002), p.32. I have not formally derived conditions under which this is sure to converge. Nonetheless, I note that this method seems to work very generally, and indeed Miranda and Fackler note that "Function iteration...often converges even when the sufficiency conditions are not met".

I found that the performance of the algorithm improved when I replaced the maximum operator in the attention function with the "smoothed maximum", described below.

Smoothed Maximum

Here I explain the "smoothed max" function which I found allowed for better performance of the fixed point iteration for finding attention equilibrium. One replaces the maximum in the attention function with the following differentiable approximation:[42]

$$\frac{m_d exp\{\alpha m_d\} + \Lambda exp\{\alpha\Lambda\}}{exp\{\alpha m_d\} + exp\{\alpha\Lambda\}}$$

Here, $\Lambda := 1 - \frac{1}{\lambda}$. As $\alpha \to \infty$ this converges to the maximum.

Robustness Checks

I now present the results from the specification using de-trended log real GDP per capita as an output gap measure. The results are rather similar to the baseline specification. Once again, the mean estimate for ϕ_π is well below unity for the Great Inflation period, above unity for the longer Great Moderation sub-sample, and just below unity for the post-1990 sub-sample. Again, the evidence against indeterminacy is "very strong" in all sub-samples. Two differences worth noting are: (i) the estimate for θ tends to be slightly higher than in the baseline specification, which is perhaps unsurprising given that the output gap series is more volatile than in the baseline specification, and (ii) the estimate for ϕ_x is somewhat lower, and more in line with estimates typically obtained in the literature.

I also present a robustness check showing that the evidence in favour of determinacy is strong even halving the firm cost of attention k_f to 0.75. I believe that this demonstrates that the results are very robust to different specifications of the attention cost.

Appendix F

8.1 Comparison to RBC Model

Both the analysis of the determinacy condition and of average inflation targeting suggest that the new Keynesian model may often exhibit muliplicity under endogenous cognitive discounting when policy is passive. This arises because when policy is passive the attention choices of different agents are strategic complements, at least in certain regions of the parameter space. Notice the upwards jump in the g mapping between the determinacy and indeterminacy regions in figure 2.1, for example. This effect arises because of the discontinuity in macroeconomic dynamics between the determinate and indeterminate regions. Multiple equilibria are only able to exist in the presence of strategic complementarities; if the g mapping is monotonically decreasing, then it is straightforward to see that if an equilibrium exists, it must be unique.

Whether the new Keynesian model generates strategic complementarities or substitutabilities between agents' attention choices depends on the shock considered. The analysis also becomes somewhat more complex when one allows for firms' expectations in the Phillips curve, and suppose that firms also optimally choose a cognitive discount factor.

One useful comparison, however, to show the importance of strategic complementarities and substitutabiliites in generating multiple equilibria, is an RBC model. Here I take a log-linearized RBC model and solve for the g mapping. I take a simple setup in which consumers save by investing in capital, which

[42] This particular formulation of the smoothed maximum is from Lange et al. (2014).

Table 9: Posterior Distribution - Alternative Output Gap Specification

	Great Inflation (1960:I to 1979:II)				Great Moderation (a) (1984:I to 2007:IV)				Great Moderation (b) (1990:I to 2007:IV)			
	Determinacy		Indeterminacy		Determinacy		Indeterminacy		Determinacy		Indeterminacy	
Parameter	Mean	90-pct. interval	Mean	90-pct. interval	Mean	90-pct. interval	Mean	90-pct. interval	Mean	90-pct. interval	Mean	90-pct. interval
ϕ_π	0.68	[0.52,0.84]	0.47	[0.32,0.59]	1.49	[0.90,2.14]	0.18	[0.08,0.30]	0.93	[0.43,1.51]	0.18	[0.08,0.31]
ϕ_x	0.30	[0.21,0.42]	0.27	[0.19,0.37]	0.47	[0.33,0.63]	0.28	[0.05,0.45]	0.49	[0.35,0.67]	0.37	[0.26,0.48]
θ	0.92	[0.88,0.95]	0.89	[0.87,0.92]	0.92	[0.90,0.94]	0.90	[0.87,0.92]	0.92	[0.90,0.94]	0.89	[0.87,0.91]
γ	2.04	[1.41,2.79]	2.16	[1.53,2.97]	2.71	[2.00,3.53]	3.02	[2.18,3.84]	2.46	[1.77,3.28]	2.80	[2.08,3.60]
ρ_η	0.75	[0.67,0.82]	0.85	[0.80,0.90]	0.63	[0.54,0.72]	0.83	[0.59,0.91]	0.61	[0.50,0.70]	0.87	[0.82,0.91]
ρ_z	0.77	[0.70,0.85]	0.72	[0.64,0.80]	0.89	[0.84,0.92]	0.84	[0.76,0.91]	0.87	[0.81,0.92]	0.81	[0.74,0.88]
ρ_i	0.71	[0.62,0.79]	0.70	[0.61,0.78]	0.86	[0.83,0.90]	0.87	[0.82,0.93]	0.87	[0.83,0.91]	0.84	[0.80,0.88]
σ_η^ε	0.11	[0.07,0.15]	0.08	[0.05,0.12]	0.08	[0.06,0.10]	0.05	[0.03,0.10]	0.08	[0.06,0.10]	0.04	[0.03,0.06]
σ_z^ε	0.84	[0.59,1.17]	1.04	[0.71,1.49]	0.43	[0.32,0.56]	0.57	[0.36,0.78]	0.44	[0.31,0.59]	0.62	[0.43,0.88]
σ_v^ε	0.16	[0.14,0.19]	0.16	[0.14,0.18]	0.12	[0.10,0.13]	0.13	[0.11,0.15]	0.10	[0.08,0.11]	0.10	[0.09,0.11]
σ_ζ			0.13	[0.06,0.23]			0.10	[0.05,0.17]			0.10	[0.05,0.17]
m_f	0.97	[0.94,0.99]	0.98	[0.98,0.99]	0.86	[0.85,0.91]	0.97	[0.95,0.99]	0.86	[0.85,0.89]	0.97	[0.96,0.98]
m_c	0.85	[0.85,0.85]	0.85	[0.85,0.85]	0.85	[0.85,0.85]	0.85	[0.85,0.85]	0.85	[0.85,0.85]	0.85	[0.85,0.85]
Log ML	-128.4		-141.7		15.6		-42.3		17.3		-25.4	

Table 10: Posterior Distribution - Low Attention Cost Specification ($k^f = 0.75$)

	Great Inflation (1960:I to 1979:II)				Great Moderation (a) (1984:I to 2007:IV)				Great Moderation (b) (1990:I to 2007:IV)			
	Determinacy		Indeterminacy		Determinacy		Indeterminacy		Determinacy		Indeterminacy	
Parameter	Mean	90-pct. interval	Mean	90-pct. interval	Mean	90-pct. interval	Mean	90-pct. interval	Mean	90-pct. interval	Mean	90-pct. interval
ϕ_π	0.65	[0.54,0.75]	0.57	[0.46,0.66]	1.61	[1.04,2.28]	0.26	[0.12,0.41]	0.97	[0.51,1.53]	0.24	[0.12,0.38]
ϕ_x	0.52	[0.40,0.66]	0.49	[0.38,0.62]	0.74	[0.53,0.97]	0.66	[0.47,0.85]	0.70	[0.53,0.90]	0.58	[0.42,0.73]
θ	0.89	[0.84,0.93]	0.87	[0.83,0.90]	0.90	[0.86,0.94]	0.90	[0.87,0.92]	0.91	[0.87,0.94]	0.89	[0.86,0.91]
γ	2.93	[2.15,3.81]	3.10	[2.29,4.06]	3.84	[2.94,4.88]	4.61	[3.66,5.70]	3.33	[2.49,4.28]	3.82	[2.95,4.77]
ρ_η	0.75	[0.67,0.81]	0.77	[0.69,0.84]	0.64	[0.54,0.73]	0.80	[0.72,0.86]	0.60	[0.50,0.70]	0.78	[0.69,0.85]
ρ_z	0.80	[0.73,0.86]	0.78	[0.70,0.85]	0.90	[0.87,0.93]	0.89	[0.84,0.92]	0.88	[0.84,0.92]	0.87	[0.82,0.92]
ρ_i	0.56	[0.46,0.66]	0.55	[0.45,0.65]	0.84	[0.80,0.87]	0.84	[0.80,0.87]	0.83	[0.78,0.87]	0.81	[0.76,0.85]
σ_η^ε	0.10	[0.07,0.14]	0.10	[0.07,0.014]	0.07	[0.05,0.09]	0.05	[0.03,0.07]	0.08	[0.06,0.10]	0.05	[0.04,0.08]
σ_z^ε	0.49	[0.37,0.64]	0.53	[0.39,0.71]	0.28	[0.22,0.35]	0.35	[0.27,0.45]	0.29	[0.22,0.37]	0.34	[0.25,0.45]
σ_v^ε	0.16	[0.14,0.19]	0.16	[0.14,0.19]	0.12	[0.11,0.14]	0.12	[0.11,0.14]	0.10	[0.09,0.12]	0.10	[0.09,0.12]
σ_ζ			0.12	[0.06,0.20]			0.08	[0.05,0.12]			0.08	[0.05,0.13]
m^f	0.98	[0.97,1.00]	0.98	[0.97,0.99]	0.92	[0.85,0.97]	0.98	[0.97,0.99]	0.91	[0.85,0.96]	0.98	[0.97,0.99]
m^c	0.85	[0.85,0.85]	0.85	[0.85,0.85]	0.85	[0.85,0.85]	0.85	[0.85,0.85]	0.85	[0.85,0.85]	0.85	[0.85,0.85]
Log ML		-53.9		-57.2		82.4		43.7		62.3		35.9

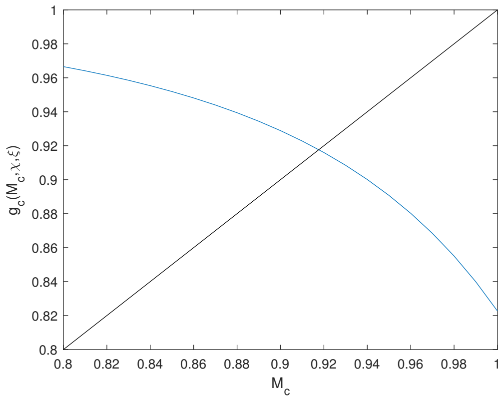


Figure 8.1: The g mapping in the RBC model

they then rent to firms. The consumer problem is:

$$\max_{\{C_{t+h},N_{t+h},K_{t+h+1}\}_{h=0}^{\infty}} \tilde{E}_t \sum_{h\geq 0} \beta^h \left(\frac{C_{t+h}^{1-\gamma}}{1-\gamma} - \frac{N_{t+h}^{1+\phi}}{1+\phi} \right) \tag{47}$$

$$\text{subject to} \tag{48}$$

$$C_{t+h} + K_{t+h+1} = (1 + r_{t+h} - \delta)K_{t+h} + w_{t+h}N_{t+h} \tag{49}$$

Here, $\tilde{E}_t$ is the subjective (myopic) expectation. Firms are perfectly competitive, and have a Cobb-Douglas production function.

$$Y_{t+h} = A_{t+h}K_{t+h}^{\alpha}N_{t+h}^{1-\alpha} \tag{50}$$

Capital and labour are then paid their marginal products:

$$w_{t+h} = (1-\alpha)A_{t+h}K_{t+h}^{\alpha}N_{t+h}^{-\alpha} \tag{51}$$

$$r_{t+h} = (1-\alpha)A_{t+h}K_{t+h}^{\alpha-1}N_{t+h}^{1-\alpha} \tag{52}$$

I assume that productivity follows are AR(1) process:

$$\log A_{t+h} = \rho \log A_{t+h-1} + \varepsilon_{t+h} \tag{53}$$

Where ε_{t+h} is independently and identically distributed. I log-linearize and solve the model. The attention choice comes from the log-linearized household Euler equation, which gives:

$$\hat{c}_t = -\gamma^{-1}(1+\beta\delta) \sum_{h\geq 1} m_c^h E_t \hat{r}_{t+h} \tag{54}$$

As above, I take the expectation of the square of the derivative with respect to attention, and use that to calculate the g mapping.

Figure 8.1 shows the results. As before M_c denotes the aggregate attention level. Notice that in the RBC model, the g mapping slopes downwards. Why does this occur? Suppose there is a positive technology shock, so that expected real interest rates rise. The higher is attention, the more that consumption decreases in response to the shock, and the more investment rises. This rise in investment attenuates the initial increase in the real interest rate. So, the higher is attention, the lower is real interest rate volatility. The lower is real interest rate volatility, the lower are incentives to pay attention. As such, agents' choices of attention are strategic substitutes. That means that the equilibrium level of attention must be unique.