Essays on Monetary Theory

by

Craig Fratrik

Department of Economics
Duke University

Date: __________________________
Approved:

____
Attila Ambrus, Advisor

____
Arjada Bardhi

____
Kyle Jurado

____
Andrea Lanteri

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2023

Abstract

Essays on Monetary Theory

by

Craig Fratrik

Department of Economics
Duke University

Date: ______________________

Approved:

____
Attila Ambrus, Advisor

____
Arjada Bardhi

____
Kyle Jurado

____
Andrea Lanteri

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics in the Graduate School of Duke University
2023

Abstract

This dissertation consists of two essays about macroeconomic theory of monetary policy. The first essay derives a general algorithm for finding optimal commitment policy when the policymaker's decision stabilizes the economy as well as informing the private sector about fundamental shocks. The paper describes three equivalent formulations of the problem facing the policymaker. The last formulation is recursive, facilitating the finding of the steady state. This paper finds the steady state for a New Keynesian central bank who has both a transitory and a persistent preference shock. Under discretion, the private sector's expected inflation is positive and persistent, limiting the ability of the central bank to achieve its output target. In contrast, under commitment, for both persistent and transient shocks, the central bank achieves negative expected inflation, allowing lower realized inflation and realized output closer to target.

The second essay introduces a new model for intermediate behavior between discretion and commitment. Instead of commitment as the ability bind a future self, this essay reframes it as how much does the current policymaker incorporate the perspective of its past self. In a monetary New Keynesian context, I define Scaled Commitment, where prior Lagrange multipliers are discounted, nesting both discretion and commitment. It also allows different degrees of commitment for the always-binding Phillips curve and occasionally-binding zero lower bound.

Dedication

To Oliver, Kai, Violet, and Ezekiel.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to acknowledge my committee members Attila Ambrus, Arjada Bardhi, Kyle Jurado, and Andrea Lanteri. Kyle, especially, I would like to thank for the support and feedback over the years. Attila, I would like to thank for supporting me when I chose to change directions part way through my PhD.

Among my colleagues, I would like to thank Jackson Bunting, Peter Horvath, Taehoon Kim, and Basil Halperin. I would especially like to thank David Min, who was an indefatigable sounding board. I would also like to thank parents and sister, Mark, Cathy, and Nicole, who have provided a great deal of support over the years. Finally, I would especially like to thank my wife, Amanda. It has been a bumpy road, and I was always glad you were on my team.

1

Introduction

This dissertation presents two chapters that expand the possible modeling of central banks, and macroeconomic policymakers in general. The first chapter considers a central bank in a linear system with quadratic losses, where the central bank has an informational advantage over the private sector. This means that the central bank's choices have a signaling role in addition to the standard stabilization role. Mertens (2016) has solved a similar models if the central bank is discretionary. This chapter solves for if the central bank has commitment, which introduces the significant complication that private-sector updating becomes a control variable.

To solve the model with an informational advantage and commitment, I present three novel formulations of the problem, and prove their equivalence. The last formulation is recursive and enables the solving for the steady state equilibrium. I solve repeat the application in Mertens (2016) for commitment instead of discretion and describe the results.

The second chapter introduces a new model of behavior between discretion and commitment. In standard models, a policymaker is committed if they can bind their future self. They are discretionary if they have no control over future actions. Debortoli and Nunes (2010) introduce a concept where the commitments have a stochastic chance of being broken, after which the policymaker reoptimizes. This chapter proposes a new framing of commitment, instead of focusing on the ability to bind a future self, it models acting a committed fashion. In particular, acting in a committed fashion means doing today what past versions of the policymaker would have wanted the agent to do. When acting fully committed in this fashion, it is performing in a *timeless* manner, a la Woodford (1999). Mathematically, Marcet and Marimon (2019) introduce Recursive Contracts which solve commitment behavior in a Bellman setup by making prior Lagrange multipliers a state variable in the problem. I extend their framework to allow for policymakers that operate in a *somewhat* committed fashion, that is they may incorporate prior Lagrange multipliers, but not necessarily to the extent that a committed agent would.

In Scaled Commitment, the central bank discounts prior Lagrange multipliers by a scalar factor. This nests discretion and commitment with discount factors of 0 an 1. In a two-period context, I apply Gul and Pesendorfer (2001) and derive the temptation function that is consistent with the intermediate behavior. In a recursive context, I apply Scaled Commitment to the standard New Keynesian model, as well as one with the occasionally-binding zero lower bound (ZLB) constraint. Two constraints allow for independent discount factors. I map out the implications for outcomes and welfare of of the different discount factors with respect to the ZLB. Finally, I show how Scaled Commitment requires only a minor modification to a standard linear quadratic regulator solution method.

2

Optimal LQR Commitment with Signaling

2.1 Introduction

This paper models a fully informed policymaker that can commit to a linear policy, where he faces a forward-looking constraint based on rational, private-sector expectations. The additional complication is that the private sector has partial information. The policymaker's choices have a signaling role and may inform the private sector. Therefore, the expectations of the private sector become another control variable for the policymaker. The structure of the model is that of a linear quadratic regulator (LQR): transitions and constraints are linear, shocks are Gaussian, and losses are quadratic. However, the private sector will use Kalman updating for their beliefs, which depends non-linearly on the policymaker's choices.

There is empirical evidence suggesting the Federal Reserve has an informational advantage over the private sector. When the public observes Federal Reserve interest rate decisions, they update their beliefs about underlying shocks to the economy, see e.g. Melosi (2016) and Nakamura and Steinsson (2018). In those papers, the interest rate was set by a Taylor rule which depended on the true shock, and the partially-informed private sector updated based on the observed interest rate.

This paper enables the solving for optimal monetary policy when a central bank with commitment has an informational advantage. Prior work has presented general solution methods for when the private sector has the same limited information or is at an informational advantage over the central bank (Svensson and Woodford (2003, 2004) respectively). When the central bank does not have an informational advantage, its policy choices do not carry any information. Therefore, in the Svensson and Woodford papers, a change in policy had linear effects and quadratic losses. They use the first-order condition to show how the optimal policy depends linearly on the fundamental state and a co-state of Lagrange multipliers from the constraints.

When an informational advantage for the central bank is present, Mertens (2016) shows how to solve for optimal policy if the central bank is discretionary. The discretionary central bank is aware that the private sector will use a Kalman gain based on its policy to update their beliefs about the state of the economy. However, the Kalman gain is linear like the rest of the model. When finding optimal discretionary policy, the central bank takes the private-sector updating as given. So like the Svensson and Woodford papers, for the purposes of optimization, a change in policy has linear effects and quadratic losses. Therefore, the first-order condition gives the central bank's best response to any specific private-sector Kalman gain. Mertens presents an algorithm to find a Kalman gain that is consistent with the central bank's best response and rational, forward-looking expectations. Crucially though, the central bank takes the Kalman gain as given and does not alter its policy in order to shape this period's private-sector updating.

Because this paper's model has both commitment and an informational advantage for the central bank, it has a complication not present in the work cited above. In the works above, the central bank does not try to change the private-sector updating process with its policies. When it does not have an informational advantage, its choices do not affect private-sector updating. And under discretion, its choices are a best response to private-sector updating, which it takes as given. Only under commitment with an informational advantage does the private-sector updating itself become a control variable.

In order to optimally shape the private-sector updating, the policymaker must commit to a policy before any shocks are realized. The distribution of initial shocks and the committed policies will be common knowledge. The private sector will use that common knowledge to update their beliefs based the signals they observe. Once some shocks are realized, the central bank will have a different preference ranking for potential policies. For a bank with commitment, the optimal policies are those that minimize unconditional expected losses, before any shocks are realized. Because they minimize unconditional expected losses, they are chosen based on the distribution of shocks and model parameters.

So that the private sector uses a Kalman filter, as in Mertens (2016), I limit the policymaker to using linear policies. This is equivalent to choosing matrices for the weightings on different inputs into the policy before any shocks are realized.[1] There is not a simple first order condition for the matrices, because their weights have non-linear effects on the model, as the mapping from policy to Kalman filter is non-linear.

I describe three novel formulations of the problem, and prove their equivalence. The first formulation is the Full History Sequence Problem (FHSP), in which the policymaker chooses a sequence of matrices that include linear weights on the full history of shocks. This is equivalent to a fully state-contingent linear policy. As a consequence, the number of weights for period t is proportional to t. The second formulation is the Commitment-Error Sequence Problem (CESP). In it, the policymaker chooses a sequence of matrices of fixed size, basing the policy each period on

[1] If the weights are chosen after any shocks are known, they could vary based on those shocks, which would mean the final policy is not linear in the inputs.

a commitment of predictable policy from past period, and a linear weighting on the private sector's prediction errors. The bulk of the proof is the equivalence between the FHSP and the CESP. It is challenging because of the dimension reduction. In the FHSP, the policymaker chooses a growing number of parameters each period, and in the CESP, the policymaker chooses a fixed number of parameters each period.

Both the FHSP and CESP choose an infinite sequence of matrices based on model parameters and the distribution of the initial shock. I call this perspective the *unconditional perspective*. Once the policymaker knows any single realized shock, it is likely that they would choose different matrices as inputs into the policy, which break the linearity of the final policy. Therefore, in all formulations the policymaker must choose the matrices before any shocks are realized. This informational posture is different than other commitment models, where the policymaker can base their policy on at least some realized shocks.[2]

The Variance Bellman Problem (VBP) is the third formulation and most tractable computationally. In it, each period there will be an augmented state composed of three parts: this period's fundamental state, the previous period's private-sector prediction of this period's fundamental state, and the previous period's private-sector prediction of this period's policy choice. When the policy is implemented, the final part will constitute the policymaker's commitment from the previous period. Like the sequence formulations, the VBP chooses matrices from the unconditional perspective. Therefore, the VBP recursive value function has a different meaning than normal Bellman or Recursive Contract value functions. In those problems, the value function represents total discounted utility from the perspective of an agent at a particular period t. In the VBP, the value function represents the expected losses

[2] See e.g. Marcet and Marimon (2019, p. 3 (1591)) where state-contingent problems are based on initial state x_0 and initial shock s_0; Svensson and Woodford (2004, p. 14 (674)) where the instrument i_t is a function of the policymaker's estimate of the underlying state $X_{t|t}$ and the prior Lagrange multiplier $\Xi_{t-1|t}$. Those models could be solved based on the unconditional perspective, but it is easier to solve them based on the information available.

starting at a period t from the *unconditional perspective.*

The value function equals unconditional expected losses starting from a certain period. Its argument is the period's augmented state covariance matrix, i.e. the unconditional distribution of the augmented state for a specific period. From that information the policymaker chooses one period of the CESP matrices. The covariance matrix and the matrices are sufficient to calculate expected losses and the next period's augmented state covariance matrix.

The novelty of this approach is that the recursion is from the unconditional perspective, i.e. is always about unconditional expected losses and distributions. It takes place before any realizations. It is a recursive formulation for the policymaker in the same position as the CESP, committing to an infinite sequence of matrices from an unconditional perspective. It is not equivalent to a sequence of policymakers through time. Once the VBP is solved, its policy functions represent a sequence of matrices that can be combined with shocks and augmented states to determine the realized policies through time. With the VBP it is possible to find the steady state of the system, in which the augmented state has the same distribution across periods. In this paper, I do so for the commitment version of the application in Mertens (2016).

There are some similarities between this paper and Marcet and Marimon (2019), in which they prove an equivalence between state-contingent commitment policies and their Recursive Contracts formulation. In their recursive formulation, the utility function of the policymaker next period is updated to incorporate how the policies next period affect prior constraints. The optimal choice of the modified utility function is the one that past policymakers would have committed themselves to. My solution differs in that I solve for policies every period based on *distributions* of states, not *realized* values. This comes from the different informational structure of my problem. By limiting myself to one-period forward constraints and linear Gaussian systems, I can embed the commitment directly into the augmented state, and therefore its distribution. But broadly, there is the similarity of drawing an equivalence between a state-contingent formulation and a recursive formulation.

The VBP is best thought of as alternative path to calculating CESP losses and determining optimal matrices. Instead of choosing all the matrices simultaneously to minimize the expectation of the entire sum of losses, the value function is used to encapsulate expected losses after this period. Then, inside the value function the minimization chooses a specific period's matrices given its augmented state distribution. This period's augmented state distribution can be combined with the matrix choices to calculate expected losses this period and the next period's augmented state distribution. The optimal choices balance effects on the expected losses this period future expected losses via the value function called with the next period's augmented state distribution. Like the CESP but unlike value functions in other models, its evaluation is entirely about expected losses and from the unconditional perspective.

This paper proceeds as follows, Section 2.2 presents a two-period version of the New Keynesian central bank model. It demonstrates the differences between the three formulations described above. Section 2.3 describes the infinite horizon model and defines formally the three formulations. Section 2.4 proves their equivalence. Section 2.5 repeats the New Keynesian model from Mertens (2016), but solving for commitment instead of discretion. Finally, Section 2.6 concludes.

2.2 Two-Period Model

This section uses the three formulations of the problem to solve a two-period New Keynesian monetary policy model. In each period, the central bank's output target is composed of two shocks, a permanent one that affects both periods, and a transitory one that affects only one period. I solve the model under both Full Information and Partial Information, to compare how the partial information changes the optimal policy of the central bank. The full information version could be solved with more traditional methods, but it is also possible to solve them based on the novel formulations of this paper.

The policymaker has a time varying target for the output gap $\overline{g}_t$[3]

$$\overline{g}_1 = v + \varepsilon_1$$

$$\overline{g}_2 = v + \varepsilon_2$$

$$\begin{bmatrix} v \\ \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma_v^2 & 0 & 0 \\ 0 & \sigma_\varepsilon^2 & 0 \\ 0 & 0 & \sigma_\varepsilon^2 \end{bmatrix}\right)$$

where v is the preference shock that persists across the two periods, and ε_t is the transitory shock. The losses take the form

$$L_t = E\left\{\pi_t^2 + (g_t - \overline{g}_t)^2\right\}$$

$$L = L_1 + L_2$$

where π_t is inflation. With two periods, I remove discounting.

The standard New-Keynesian Phillips Curve is $\pi_t = \kappa g_t + \beta E\{\pi_{t+1}\}$. For simplicity, I use $\kappa = \beta = 1$ and $\pi_3 = 0$. There are two constraints,

$$\pi_1 = g_1 + \pi_{2|1} \tag{2.1}$$

$$\pi_2 = g_2 \tag{2.2}$$

Forward-looking expectations are taken based on the private sector's information set, $\pi_{2|1} = E\{\pi_2 | I_1^{ps}\}$. Under Full Information, $I_1^{ps} = \{v, \varepsilon_1\}$, and under Partial Information, $I_1^{ps} = \{\pi_1\}$. For all problem formulations, the private sector will use a Kalman update to estimate v and ε_1 based on either (v, ε_1) or π_1

$$E\left\{\begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix} | I_1^{ps}\right\} = \begin{bmatrix} v_{|1} \\ \varepsilon_{1|1} \end{bmatrix} = K^{\{FI,PI\}} \begin{bmatrix} v \\ \varepsilon_1 \\ \pi_1 \end{bmatrix}$$

[3] I use unconventional variable g for the output gap, because x and y are used in the proof.

with $K^{\cdot}$ maps (v, ε_1, π) to private sector beliefs about the shocks. Under Full Information, the private sector directly observes the shocks, and under Partial Information, there will linear weights based on π_1

$$K^{FI} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{2.3}$$

$$\begin{bmatrix} v_{|1} \\ \varepsilon_{1|1} \end{bmatrix} = K^{PI,\pi} \pi_1$$

$$K^{PI} = \begin{bmatrix} 0 & 0 & K^{PI,\pi} \end{bmatrix} \tag{2.4}$$

In all three formulations, π_1 will depend linearly on v and ε_1. Let row vector G_1 hold those linear weights

$$\pi_1 = G_1 \begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix}$$

and let the shocks have distribution $\Sigma^{v\varepsilon}$

$$\begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix} \sim N\left(0, \Sigma^{v\varepsilon}\right)$$

$$\Sigma^{v\varepsilon} = \begin{bmatrix} \sigma_v^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{bmatrix}$$

The Kalman update for Partial Information is given by

$$K^{PI,\pi} = \Sigma^{v\varepsilon} G_1^T \left(G_1 \Sigma^{v\varepsilon} G_1^T \right)^{-1}$$

Controlling $K^{PI,\pi}$ is the reason that the central bank commits to G_1 before any shocks realize.

A key to my approach is that for commitment, the optimal linear policy can be chosen before the first shock realizes, based on covariances and minimizing unconditional losses. In my proof, I prove the equivalence of the three problem formulations, which I will apply here. The policymaker is a Ramsey planner and chooses all instruments and forward-looking outcomes subject to the Forward-Looking Constraint. In this problem that would mean choosing both π_t and g_t such that (2.1) and (2.2) are met. However, we can simplify the problem by letting the policymaker optimize the choice of π_t, and then get the implied g_t: $g_1 = \pi_1 - \pi_{2|1}$ and $g_2 = \pi_2$.

2.2.1 Summary of Problem Formulations

Here is the central bank's problem as a Full History Sequence Problem (FHSP)

$$
\begin{aligned}
V\left(\Sigma^{v\varepsilon}\right) &= \min_{G_1,G_2} E\left\{L_1 + L_2\right\} \\
\text{s.t.} \quad \begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix} &\sim N\left(0, \Sigma^{v\varepsilon}\right) \\
w_2 &\sim N\left(0,1\right) \\
\begin{bmatrix} v \\ \varepsilon_2 \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix} + \begin{bmatrix} 0 \\ \sigma_\varepsilon \end{bmatrix} w_2 \\
\pi_1 &= G_1 \begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix} \\
\pi_2 &= G_2 \begin{bmatrix} v \\ \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \\
g_1 &= \pi_1 - \pi_{2|1} \\
g_2 &= \pi_2
\end{aligned}
$$

where unconditional expected losses will depend on the distribution of shocks, $\Sigma^{v\varepsilon}$, the variance of ε_2 is controlled by the σ_ε multiplied by unit variance w_2, G_1 is a row vector of length 2, G_2 is a row vector of length 3, and $\pi_{2|1} = E\left\{\pi_2 | I_1^{ps}\right\}$ which can vary across informational setups. In particular, under both Full and Partial information, $\pi_{2|1}$ will depend on G_2. Under Partial Information, it also depends on the information in π_1,

$$\begin{bmatrix} v_{|1} \\ \varepsilon_{1|1} \end{bmatrix} = K^{\{FI,PI\}} \begin{bmatrix} v \\ \varepsilon_1 \\ \pi_1 \end{bmatrix} = K^{\{FI,PI\}} \begin{bmatrix} I \\ G_1 \end{bmatrix} \begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix}$$

$$\pi_{2|1} = G_2 \begin{bmatrix} v_{|1} \\ \varepsilon_{1|1} \\ 0 \end{bmatrix} = G_2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} K^{\{FI,PI\}} \begin{bmatrix} I \\ G_1 \end{bmatrix} \begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix}$$

where K^{FI} and K^{PI} are defined in (2.3) and (2.4). Using the matrices to choose π_t, and then calculating the consistent g_t means that all possible G_1, G_2 meet the Forward-Looking Constraint.

In the FHSP, the policymaker is minimizing the total sum of losses. The choice variables are two row vectors that determine how the policies depend on the full history up to that period.

Here is the central bank's problem as a Commitment Error Sequence Problem

(CESP)

$$
\begin{aligned}
W\left(\Sigma^{v\varepsilon}\right) &= \min_{G_1^e, G_1^c, G_2^e} E\left\{L_1 + L_2\right\} \\
\text{s.t.} \quad \begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix} &\sim N\left(0, \Sigma^{v\varepsilon}\right) \\
w_2 &\sim N\left(0, 1\right) \\
\begin{bmatrix} v \\ \varepsilon_2 \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix} + \begin{bmatrix} 0 \\ \sigma_\varepsilon \end{bmatrix} w_2 \\
\pi_1 &= G_1^e \begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix} \\
\begin{bmatrix} v_{|1} \\ \varepsilon_{1|1} \end{bmatrix} &= K^{\{FH,PI\}} \begin{bmatrix} v \\ \varepsilon_1 \\ \pi_1 \end{bmatrix} \\
\pi_{2|1} &= G_1^c \begin{bmatrix} v_{|1} \\ \varepsilon_{1|1} \end{bmatrix} \\
\pi_2 &= \pi_{2|1} + G_2^e \left(\begin{bmatrix} v \\ \varepsilon_2 \end{bmatrix} - \begin{bmatrix} v_{|1} \\ 0 \end{bmatrix} \right) \\
g_1 &= \pi_1 - \pi_{2|1} \\
g_2 &= \pi_2
\end{aligned}
$$

where again unconditional losses depend on the distribution of shocks $\Sigma^{v\varepsilon}$; row vectors G_1^e, G_1^c, and G_2^e are all of length 2.[4] The superscript e signifies error for the private sector's prediction error. The superscript c in G_1^c stands for the commitment the policymaker makes regarding $\pi_{2|1}$. Finally, by using the matrices to determine π_t, and then getting the implied g_t, all possible $\{G_1^e, G_1^c, G_2^e\}$ meet the Forward-Looking Constraint.

The CESP structures the problem differently than the FHSP. In the FHSP, the

[4] In the full proof, I would allow G_2^c to be multiplied by $\begin{bmatrix} v_{|1} \\ \varepsilon_{1|1} \\ \pi_{1|1} \end{bmatrix}$, but in this problem $\pi_{1|1} \in \operatorname{span}\left(\begin{bmatrix} v_{|1} \\ \varepsilon_{1|1} \end{bmatrix}\right)$, so I simplified.

expected behavior of $\pi_{2|1}$ was determined by the information the private sector had, I_1^{ps}, and the policy G_2. Here π_2 is decomposed into the private-sector predictable part, which is controlled by G_1^c, and the part that depends on private-sector prediction error, G_2^e. Note that by construction, whatever the choice of G_2^e,

$$E\left\{G_1^c \begin{bmatrix} v_{|1} \\ \varepsilon_{1|1} \end{bmatrix} |I_1^{ps}\right\} = G_1^c \begin{bmatrix} v_{|1} \\ \varepsilon_{1|1} \end{bmatrix}$$

$$E\left\{\begin{bmatrix} v \\ \varepsilon_2 \end{bmatrix} - \begin{bmatrix} v_{|1} \\ 0 \end{bmatrix} |I_1^{ps}\right\} = 0$$

$$\text{and therefore, } E\left\{\pi_2 | I_1^{ps}\right\} = G_1^c \begin{bmatrix} v_{|1} \\ \varepsilon_{1|1} \end{bmatrix}$$

In other words, π_2 is chosen in such a way as it is guaranteed to be consistent with the rational expectation which was set by $K^{\{FI,PI\}}$ and G_1^c.

Now we will consider the Variance Bellman Problem (VBP), which differs from traditional Bellman problems. The normal pattern for a Bellman function is that the recursive call captures the discounted losses of the agent *in the next period.* For example, if we were considering a consumption decision based on a wealth state,

$$V(w) = \max_c u(c) + \beta E\left\{V(w')\right\}$$

$V(w')$ would represent the discounted utility of the consumer in the next period. It's possible to use $V(w_0)$ above to calculate a state contingent sequence $\{c_t\}_{t=0}^{\infty}$, but it is equally valid to think about each c_t being decisions made *at different times.*

In contrast, the recursive function in a VBP does not accept as argument a state variable, but rather a covariance representing an augmented-state distribution. In the first period, the augmented state will be

$$y_1 = \begin{bmatrix} v \\ \varepsilon_1 \\ v_{|0} \\ \varepsilon_{1|0} \\ \pi_{1|0} \end{bmatrix} = \begin{bmatrix} v \\ \varepsilon_1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \sim N\left(0, \Sigma_1^y\right)$$

$$\begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix} = e^{v\varepsilon} y_1$$

$$e^{v\varepsilon} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

The (mean-zero) augmented state y_1 has five elements, but in the first period the private sector has no information or expectation coming into the period, so the bottom three elements of y_1 will be uniformly 0, as will their covariances,

$$\Sigma_1^y = \begin{bmatrix} \Sigma^{v\varepsilon} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Both Marcet and Marimon (2019) and Mertens (2016) have recursive value functions with augmented state arguments. In the former, the state is augmented with Lagrange multipliers representing prior commitments, and in the latter, the state is augmented with the private sector's estimate of the state coming into the period. For this problem the Mertens (2016) state would be $\begin{bmatrix} v & \varepsilon_1 & v_{|0} & \varepsilon_{1|0} \end{bmatrix}'$. In the VBP, the argument is Σ_1^y instead of y_1. The central bank will commit to policies for period 1 based on the distribution Σ_1^y, instead of choosing policies after observing any part of y_1. (Note that because of the finite horizon, the value function and choices will be different for the first period and second period. Both, however will accept unconditional distributions as arguments: Σ_1^y and Σ_2^y.)

For the two-period model the FHSP and CESP are tractable. The key feature of the VBP is that the value function for the second period, $U_2\left(\Sigma_2^y\right)$, can be calculated without reference to prior policies beyond the covariance matrix Σ^y_2. This demonstrates how the VBP is capable of transforming the infinite problem into a recursive one, and that Σ^y is a sufficient statistic for calculating optimal G^e and G^c choices.

Given the distribution of the y_1 for period 1, the policymaker can minimize his choice of G^e_1 and G^c_1, taking into account the fact that they will affect expected utility for later periods, based on changing the next period's distribution Σ^y_2. The definition of U_1 is given by

$$
\begin{aligned}
U_1\left(\Sigma^y_1\right) &= \min_{G^e_1, G^c_1} E\left\{L_1\right\} + U_2\left(\Sigma^y_2\right) \\
\text{s.t. } y_1 &\sim N\left(0, \Sigma^y_1\right) \\
\pi_1 &= G^e_1 \begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix} = G^e_1 e^{v\varepsilon} y_1 \\
\pi_{2|1} &= G^c_1 \begin{bmatrix} v_{|1} \\ \varepsilon_{1|1} \end{bmatrix} \\
g_1 &= \pi_1 - \pi_{2|1}
\end{aligned}
$$

where G^e_1 and G^c_2 are row vectors of length 2, and Σ^y_2 is calculated below.

Beliefs updates use $K^{\{FI,PI\}}$ is given by (2.3) or (2.4), depending on which informational structure we are modeling.

$$
\begin{bmatrix} v_{|1} \\ \varepsilon_{1|1} \end{bmatrix} = K^{\{FI,PI\}} \begin{bmatrix} v \\ \varepsilon_1 \\ \pi_1 \end{bmatrix} = K^{\{FI,PI\}} \begin{bmatrix} I \\ G^e_1 \end{bmatrix} e^{v\varepsilon} y_1
$$

The system evolves according to

$$
\begin{aligned}
w_2 &\sim N\left(0, 1\right) \\
\begin{bmatrix} v \\ \varepsilon_2 \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix} + \begin{bmatrix} 0 \\ \sigma_\varepsilon \end{bmatrix} w_2
\end{aligned}
$$

Now I show how Σ^y_2 is calculated from Σ^y_1, G^e_1, and G^c_1. Note first that

$$
\begin{bmatrix} v_{|1} \\ \varepsilon_{2|1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_{|1} \\ \varepsilon_{1|1} \end{bmatrix}
$$

so together

$$\begin{bmatrix} v_{|1} \\ \varepsilon_{2|1} \\ \pi_{2|1} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \\ G_1^c \end{bmatrix} \begin{bmatrix} v_{|1} \\ \varepsilon_{1|1} \end{bmatrix}$$

Therefore we can calculate the entire transition matrices A_1^y and B^y,

$$y_2 = \begin{bmatrix} v \\ \varepsilon_2 \\ v_{|1} \\ \varepsilon_{2|1} \\ \pi_{2|1} \end{bmatrix} \sim N\left(0, \Sigma_2^y\right)$$

$$= A_1^y y_1 + B^y w_2$$

$$w_2 \sim N\left(0, 1\right)$$

$$A_1^y = \begin{bmatrix} & \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} e^{v\varepsilon} \\ \begin{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \\ G_1^c \end{bmatrix} K^{\{FI,PI\}} \begin{bmatrix} I \\ G_1^e \end{bmatrix} e^{v\varepsilon} \end{bmatrix}$$

$$B^y = \begin{bmatrix} 0 \\ \sigma_\varepsilon \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

and from these combined we can calculate the distribution of y_2

$$\Sigma_2^y = A_1^y \Sigma_1^y \left(A_1^y\right)^T + B^y \left(B^y\right)^T$$

As a reminder, U_1 and U_2 take as arguments the covariance matrices of augmented states y_1 and y_2. Note especially that part of y_2 is $\pi_{2|1}$, which is a choice variable *of period 1 optimization.* $\pi_2\left(y_2, \Sigma_2^y\right)$ will be linear in y_2 for all Σ_2^y , and it will be constructed in such a way to ensure that prior period's private-sector expectation is followed through upon, $E\left\{\pi_2\left(y_2\right)|I_1^{ps}\right\} = \pi_{2|1}$. U_2 includes the optimization of period 2 policies, but those policies always ensure that

$$\pi_{2|1} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \end{bmatrix} y_2$$

is a rational expectation. The choices that remain for the optimization of period 2 policies are only things that do not change $\pi_{2|1}$.

The second period covariance has the form

$$y_2 = \begin{bmatrix} v \\ \varepsilon_2 \\ v_{|1} \\ \varepsilon_{2|1} \\ \pi_{2|1} \end{bmatrix} = \begin{bmatrix} v \\ \varepsilon_2 \\ v_{|1} \\ 0 \\ \pi_{2|1} \end{bmatrix} \sim N\left(0, \Sigma_1^y\right)$$

$$e^{v\varepsilon} \equiv \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} v \\ \varepsilon_2 \end{bmatrix} = e^{v\varepsilon} y_2$$

$$e_{2|1}^{v\varepsilon} \equiv \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} v_{|1} \\ \varepsilon_{2|1} \end{bmatrix} = \begin{bmatrix} v_{|1} \\ 0 \end{bmatrix} = e_{2|1}^{v\varepsilon} y_2$$

$$e_{2|1}^{\pi} \equiv \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\pi_{2|1} = e_{2|1}^{\pi} y_2$$

Here is the period 2 value function

$$U_2\left(\Sigma_2^y\right) = \min_{G_2^e} E\left\{L_2\right\}$$

$$\text{s.t. } y_2 \sim N\left(0, \Sigma_2^y\right)$$

$$\pi_2 = \pi_{2|1} + G_2^e \left(\begin{bmatrix} v \\ \varepsilon_2 \end{bmatrix} - \begin{bmatrix} v_{|1} \\ \varepsilon_{2|1} \end{bmatrix} \right)$$

$$= \left(e_{2|1}^{\pi} + G_2^e \left(e^{v\varepsilon} - e_{2|1}^{v\varepsilon}\right)\right) y_2$$

$$g_2 = \pi_2$$

so the optimization of period 2 policies amounts to choosing G_2^e that minimizes losses, given the distribution of y_2.[5]

[5] In this particular case no matter Σ_2^y and informational setup, the optimal $G_2^e = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$.

The usual Bellman equation describes the problem that a policymaker will face for some period, and how they will choose the optimal policy when that period arrives. It is from the perspective of that period. Recursive Contracts can be thought of as solving the problem from the original perspective, after x_0 realizes. Alternatively it can be thought of as if the policymaker's utility function is modified in the specified manner, then when period t arrives, he will choose to act in the way that he would have committed himself to act in that period.

The VBR formulation is different than both of these. Optimal commitment with signaling must be determined before any shocks are realized, from the unconditional perspective. The VBR and its equivalence shows that optimal policy for period t can be calculated based only on the covariance the augmented state for period t, which has a fixed dimension of $(2N_x + N_a) \times (2N_x + N_a)$. The optimal policy are some matrices that will be combined with the augmented state in order to form the period t policy. But these matrices must be chosen from the unconditional perspective. Unlike the Bellman equation, and unlike the second interpretation for Recursive Contracts, we cannot think of what a policymaker will choose to do once t arrives. As soon as any shocks have been realized, the weights the policymaker would choose are changed.

Summarizing the three approaches. The FHSP chooses row vectors, G_1 and G_2 size 2 and 3 respectively, which are the weights for the state-contingent policy of the full history. This is equivalent to choosing the linear functions $\pi_1(v, \varepsilon_1)$ and $\pi_2(v, \varepsilon_1, \varepsilon_2)$.

$$\pi_1 = G_1 \begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix}$$

$$\pi_2 = G_2 \begin{bmatrix} v \\ \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

Under Full Information, these weights could be chosen after v and ε_1 are realized, however under Partial Information, they must be committed to in advance as they affect the information updating done by the private sector in the first period.

The CESP chooses three row vectors: G_1^e, G_1^c, and G_2^e all of length 2, where

$$
\begin{aligned}
\pi_1 &= G_1^e \begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix} \\
\pi_{2|1} &= G_1^c \begin{bmatrix} v_{|1} \\ \varepsilon_{1|1} \end{bmatrix} \\
&= G_1^c K^{\{FH,PI\}} \begin{bmatrix} I \\ G_1^e \end{bmatrix} \begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix} \\
\pi_2 &= \pi_{2|1} + G_2^e \left(\begin{bmatrix} v \\ \varepsilon_2 \end{bmatrix} - \begin{bmatrix} v_{|1} \\ 0 \end{bmatrix} \right)
\end{aligned}
$$

where K^{PI} is a function of $\Sigma^{v\varepsilon}$ and the weights in G_1^e. Again, under Full Information, these three row vectors could be chosen after realization of v and ε_1, but for Partial Information, they must be committed to in advance.

Finally, the VBP is choosing vector functions, $G_1^e(\Sigma_1^y)$, $G_1^c(\Sigma_1^y)$, and $G_2^e(\Sigma_2^y)$ with a crucial difference in how π_2 is calculated.

$$
\begin{aligned}
y_1 &\sim N(0, \Sigma_1^y) \\
\pi_1 &= G_1^e(\Sigma_1^y)\, e^{v\varepsilon} y_1 \\
\pi_{2|1} &= G_1^c(\Sigma_1^y) \begin{bmatrix} v_{|1} \\ \varepsilon_{1|1} \end{bmatrix} \\
&= G_1^c(\Sigma_1^y)\, K^{\{FH,PI\}} \begin{bmatrix} I \\ G_1^e \end{bmatrix} e^{v\varepsilon} y_1 \\
&\in y_2
\end{aligned}
$$

The distribution of y_2 is the argument to U_2, and its components are

$$
y_2 = \begin{bmatrix} v \\ \varepsilon_2 \\ v_{|1} \\ \varepsilon_{2|1} \\ \pi_{2|1} \end{bmatrix} \sim N(0, \Sigma_2^y)
$$

In the VBP, the optimization of period 2 policies chooses them as

$$\begin{aligned}\pi_2 &= \left(e^{\pi}_{c|p} + G^e_2\left(\Sigma^y_2\right)\left(e^{v\varepsilon} - e^{v\varepsilon}_{c|p}\right)\right) y_2 \\ &= \pi_{2|1} + G^e_2\left(\Sigma^y_2\right)\left(\begin{bmatrix} v \\ \varepsilon_2 \end{bmatrix} - \begin{bmatrix} v_{|1} \\ 0 \end{bmatrix}\right)\end{aligned}$$

The three formulations have different approaches to the problem, however they yield the same optimal policies.

2.2.2 Results

For all the formulations, the policies under Full Information do not depend on the variances of the shocks. All three policies also choose a 2-element row vector for weights determining π_1

$$G^{FI}_1 = \begin{bmatrix} 0.6 & 0.4 \end{bmatrix}$$

For the second period policy, the FHSP formulation central bank chooses

$$G^{FI}_2 = \begin{bmatrix} 0.2 & -0.2 & 0.5 \end{bmatrix}$$

Under CESP, the policymaker chooses

$$\begin{aligned} G^{c,FI}_1 &= \begin{bmatrix} 0.2 & -0.2 \end{bmatrix} \\ G^{e,FI}_2 &= \begin{bmatrix} 0.5 & 0.5 \end{bmatrix} \end{aligned}$$

although the first term in $G^{e,FI}_2$ has no effect, because $v = v_{|1}$.

To explore the value functions in the VBP precisely, I will show below how losses would be calculated for any G^e_1, G^c_1. Label the elements as

$$\begin{aligned} G^{e,FI}_1 &= \begin{bmatrix} e_v & e_c \end{bmatrix} \\ G^{c,FI}_1 &= \begin{bmatrix} c_v & c_\varepsilon \end{bmatrix} \end{aligned}$$

and let the variances be unit,

$$\begin{bmatrix} v \\ \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \sim N\left(0, I\right)$$

The VBP optimization will make the same choices as the CESP, however we can calculate U_1 and U_2 explicitly. Σ_2^y will have the form

$$\Sigma_2^y = \begin{bmatrix} 1 & 0 & 1 & 0 & c_v \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & c_v \\ 0 & 0 & 0 & 0 & 0 \\ c_v & 0 & c_v & 0 & c_v^2 + c_\varepsilon^2 \end{bmatrix}$$

Whatever the informational structure, an optimal choice is $G_2^e = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$. So solving for $U_2^{FI}\left(\Sigma_2^y\right)$

$$\pi_{2|1} = c_v v + c_\varepsilon \varepsilon_1$$

$$\pi_2 = \pi_{2|1} + \frac{1}{2}\varepsilon_2$$

$$E\left\{\pi_2^2\right\} = c_v^2 + c_\varepsilon^2 + \frac{1}{4}$$

$$g_1 - \overline{g}_1 = \pi_2 - v - \varepsilon_2$$

$$E\left\{\left(\pi_2 - v - \varepsilon_2\right)^2\right\} = \left(1 - c_v\right)^2 + c_\varepsilon^2 + \frac{1}{4}$$

$$U_2^{FI}\left(\Sigma_2^y\right) = E\left\{L_2\right\} = 2c_\varepsilon^2 + c_v^2 + \left(1 - c_v\right)^2 + \frac{1}{2}$$

From only the variables in Σ_2^y, we can give the precise formula for U_2^{FI}. This means that we can also do the same for U_1 depending only on

$$\pi_1 = e_v v + e_\varepsilon \varepsilon_1$$

$$\pi_{2|1} = c_v v + c_\varepsilon \varepsilon_1$$

$$\begin{aligned} g_1 &= \pi_1 - \pi_{2|1} \\ &= \left(e_v - c_v\right) v + \left(e_\varepsilon - c_\varepsilon\right) \varepsilon_1 \end{aligned}$$

$$g_1 - \overline{g}_1 = \left(e_v - c_v - 1\right) v + \left(e_\varepsilon - c_\varepsilon - 1\right) \varepsilon_1$$

which lets us precisely define U_1^{FI},

$$U_1^{FI}\left(\Sigma_1^y\right) = E\left\{L_1\right\} + U_2^{FI}\left(\Sigma_2^y\right)$$
$$= e_v^2 + e_\varepsilon^2 + \left(e_v - c_v - 1\right)^2 + \left(e_\varepsilon - c_\varepsilon - 1\right)^2 + U_2^{FI}\left(\Sigma_2^y\right)$$

The first order conditions give the same choices as the CESP

$$G_1^{e,FI} = \begin{bmatrix} 0.6 & 0.4 \end{bmatrix}$$
$$G_1^{c,FI} = \begin{bmatrix} 0.2 & -0.2 \end{bmatrix}$$

The forward-looking expectations increase losses for v and decrease them for ε_1. This is because v increases both $\overline{g}_1$ and $\overline{g}_2$, while ε_1 only affects $\overline{g}_1$. From the perspective of period 1, the policymaker prefers that $E\left\{\pi_{2|1}|v\right\}/v$ and $E\left\{\pi_{2|1}|\varepsilon_1\right\}/\varepsilon_1$ to be as close to -1 as possible. But purely from the perspective of period 2, the policymaker would prefer $E\left\{y_{2|1} = \pi_{2|1}|v\right\} = \frac{1}{2}v$. Balancing these two costs, the policymaker chooses $E\left\{\pi_{2|1}|v\right\} = \frac{1}{5}v$. In response to ε_1, the ideal reaction from the perspective of period 2 is to ignore it, so the balance between periods produces $E\left\{\pi_2|\varepsilon_1\right\} = -\frac{1}{5}\varepsilon_1$. This balances extra variance in π_2 and y_2 against the benefit of the expectation to period 1 losses.

Under full information, the only surprise in period 2 is ε_2, and it is independent of period 1 policies. The response to ε_2 is fully optimized so $E\left\{\pi_2|\varepsilon_2\right\} = \frac{1}{2}\varepsilon_2$, but the response to v and ε_1 are chosen to balance costs and benefits across periods.

Partial Information

Because the private sector is doing signal extraction, the numerical values will depend on the variance of the first period shocks. For the numerical results below, I give them both unit variance.

$$\begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix} \sim N\left(0, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

In the Full History formulation, the solution is

$$G_1^{PI} = \begin{bmatrix} 0.311 & 0.56 \end{bmatrix}$$

$$\pi_1 = G_1^{PI} \begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix}$$

$$G_2^{PI} = \begin{bmatrix} 0.344 & -0.28 & 0.5 \end{bmatrix}$$

$$\pi_2 = G_2^{PI} \begin{bmatrix} v \\ \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

The private sector updates their beliefs according to

$$K^{PI,\pi} = G_1^T \left(G_1 G_1^T\right)^{-1} = \begin{bmatrix} 0.76 \\ 1.37 \end{bmatrix}$$

$$\begin{bmatrix} v_{|1} \\ \varepsilon_{1|1} \end{bmatrix} = K^{PI,\pi} \pi_1$$

The central banker chooses to put less weight on v than ε_1, because it causes private sector ascribe more of the observed π_1 to the temporary shock, than the persistent shock. The net result is

$$\begin{bmatrix} v_{|1} \\ \varepsilon_{1|1} \end{bmatrix} = K^{PI,\pi} G_1 \begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix}$$

$$= \begin{bmatrix} 0.24 & 0.42 \\ 0.42 & 0.76 \end{bmatrix} \begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix}$$

After we calculate the private-sector beliefs, we can determine the implied $\pi_{2|1}$,

$$\pi_{2|1} = \begin{bmatrix} 0.344 & -0.28 \end{bmatrix} \begin{bmatrix} 0.24 & 0.42 \\ 0.42 & 0.76 \end{bmatrix} \begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix}$$

$$= \begin{bmatrix} -0.038 & -0.068 \end{bmatrix} \begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix}$$

We see that for both shocks, the private sector response of $\pi_{2|1}$ is negative. This is beneficial for first period losses because recall $g_1 = \pi_1 - \pi_{2|1}$, and the central bank wants higher g_1 and lower π_1 responses to the shocks.

Ignoring effects on first period expectations, the central bank would like $G_2 = \begin{bmatrix} 0.5 & 0 & 0.5 \end{bmatrix}$. Under Full Information it chooses, $G_2^{FI} = \begin{bmatrix} 0.2 & -0.2 & 0.5 \end{bmatrix}$. The reason the first element drops to 0.2 is because of its deleterious effect on $\pi_{2|1}$ and first period losses. Under Partial Information, the central bank chooses $G_2^{PI} = \begin{bmatrix} 0.344 & -0.28 & 0.5 \end{bmatrix}$, with higher first element, and yet $\pi_{2|1}$ still negatively responds to v shocks. This is precisely because of the information shaping in the choice of G_1^{PI}. For a shock $(v = 1, \varepsilon_1 = 0)$,

$$\begin{bmatrix} v_{|1} \\ \varepsilon_{1|1} \end{bmatrix} = \begin{bmatrix} 0.24 \\ 0.42 \end{bmatrix}$$

so the negative weight on ε_1 in G_2 causes the $\pi_{2|1} = -0.038$, reducing losses in period 1.

I numerically solve the problem using all three formulations, and they yield the same net policies, but the policymaker constructs π_2 differently. The first matrix is the same as the FHSP, $G_1^{e,PI} = G_1^{PI}$. And G_2^e is pinned down as well. However, there is more than one optimal G_1^c, because $\begin{bmatrix} v_{|1} & \varepsilon_{1|1} \end{bmatrix}'$ is collinear as it is a vector multiplied by π_1, the private sector's only information. I present below the representation value that only puts weight on $v_{|1}$. In the CESP, the central banker is choosing all the matrices simultaneously.

$$G_1^{e,PI} = \begin{bmatrix} 0.311 & 0.56 \end{bmatrix}$$

$$\pi_1 = G_1^{e,PI} \begin{bmatrix} v \\ \varepsilon_1 \end{bmatrix}$$

$$G_1^{c,PI} = \begin{bmatrix} -0.16 & 0 \end{bmatrix}$$

$$\pi_{2|1} = G_1^{c,PI} \begin{bmatrix} v_{|1} \\ \varepsilon_{1|1} \end{bmatrix}$$

$$G_2^{e,PI} = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$$

$$\pi_2 = \pi_{2|1} + G_2^{e,PI} \left(\begin{bmatrix} v \\ \varepsilon_2 \end{bmatrix} - \begin{bmatrix} v_{|1} \\ 0 \end{bmatrix} \right)$$

Again to demonstrate how the VBP works, we will calculate its losses for any choices of G_1^e, and G_1^c,

$$G_1^{e,PI} = \begin{bmatrix} e_v & e_\varepsilon \end{bmatrix}$$

$$G_1^{c,PI} = \begin{bmatrix} c_v & 0 \end{bmatrix}$$

The final quantity we will define now, and then calculate later is, $\sigma_{v|1}^2 = \text{Var}\left(v_{|1}\right)$. Σ_2^y will have the form

$$\Sigma_2^y = \begin{bmatrix} 1 & 0 & \sigma_{v|1}^2 & 0 & c_v\sigma_{v|1}^2 \\ 0 & 1 & 0 & 0 & 0 \\ \sigma_{v|1}^2 & 0 & \sigma_{v|1}^2 & 0 & c_v\sigma_{v|1}^2 \\ 0 & 0 & 0 & 0 & 0 \\ c_v\sigma_{v|1}^2 & 0 & c_v\sigma_{v|1}^2 & 0 & c_v^2\sigma_{v|1}^2 \end{bmatrix}$$

Again the optimal second period choice is always $G_2^e = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$,

$$\pi_{2|1} = c_v v_{|1}$$

$$\pi_2 = \pi_{2|1} + \frac{1}{2}\left(v - v_{|1}\right) + \frac{1}{2}\varepsilon_2$$

$$g_2 - \overline{g}_2 = \left(c_v - 1\right) v_{|1} - \frac{1}{2}\left(v - v_{|1}\right) - \frac{1}{2}\varepsilon_2$$

because $v_{|1}$ is orthogonal to $\left(v - v_{|1}\right)$, the unconditional losses are,

$$U_2^{PI}\left(\Sigma_2^y\right) = \left(c_v^2 + \left(c_v - 1\right)^2\right)\sigma_{v|1}^2 + \frac{1}{2}\left(1 - \sigma_{v|1}^2\right) + \frac{1}{2}$$

Note that because $\sigma_{v|1}^2 < 1$, and $c_v < 0.5$, these losses are increasing in $\sigma_{v|1}^2$. The better the private sector's estimate, the worse losses are in the second period. This is because the central bank must follow through on the commitment $\pi_{2|1}$ with regard to the private sector's estimate of $v_{|1}$, but can freely optimize the prediction error, $v - v_{|1}$. While $c_v < 0.5$, which will always be the case for optimal choices, the second period losses are decreasing in c_v.

Now moving back to $U_1^{PI}\left(\Sigma_1^y\right)$, again naming variables for the choices

$$G_1^{e,PI} = \begin{bmatrix} e_v & e_\varepsilon \end{bmatrix}$$

The optimal estimate for v is $v_{|1} = K_{\pi}^{v,PI}\pi_1$ with

$$\begin{aligned} K_{\pi}^{v,PI} &= \frac{e_v}{e_v^2 + e_\varepsilon^2} \\ \sigma_{v|1}^2 &= \frac{e_v^2}{e_v^2 + e_\varepsilon^2} \\ \pi_{2|1} &= c_v K_{\pi}^{v,PI}\pi_1 \\ g_1 &= \left(1 - c_v K_{\pi}^{v,PI}\right)\pi_1 \\ &= \left(1 - c_v K_{\pi}^{v,PI}\right)\left(e_v v + e_\varepsilon \varepsilon_1\right) \end{aligned}$$

Let $r_g = 1 - c_v K_{\pi}^{v,PI}$. Therefore,

$$\begin{aligned} E\left\{L_1\right\} &= e_v^2 + \left(1 - r_g e_v\right)^2 + e_\varepsilon^2 + \left(1 - r_g e_\varepsilon\right)^2 \\ U_1^{PI}\left(\Sigma_1^y\right) &= E\left\{L_1\right\} + U_2^{PI}\left(\Sigma_2^y\right) \end{aligned}$$

Because $K_{\pi}^{v,PI}$ has e_v and e_ε in the denominator, the first order condition for them gives a cubic equation, and I solve them numerically.

This section's two period exercise demonstrated two things. First it showed concretely how the three formulations address the problem. Second, in terms of outcomes, it showed how under Partial Information, the central bank has an incentive to place more relative weight on the temporary shock, in order to worsen the private sector's information. Lower quality private-sector information gives the central bank more opportunity to reoptimize in the second period.

2.3 General Model

In this section, I present the general recursive model, and the three formulation definitions. There are N_x backward-looking state variables in vector x_t, and N_a action variables available to the policymaker in vector a_t. Some macroeconomic frameworks separate the policymaker's instruments, e.g. interest rate, from forward-looking outcomes, e.g. inflation and output. I follow Marcet and Marimon (2019) in treating the policymaker as a Ramsey planner, who chooses both instruments and forward-looking outcomes subject to the constraints.

Let $N_{xa} \equiv N_x + N_a$. Each period the policymaker faces a quadratic loss in the form a positive, semi-definite matrix L of size $N_{xa} \times N_{xa}$,

$$L_t = \begin{bmatrix} x_t \\ a_t \end{bmatrix}^T L \begin{bmatrix} x_t \\ a_t \end{bmatrix} \tag{2.5}$$

In the very first period, $x_0 \sim N\left(0, \Sigma_0^x\right)$. The backward-looking state evolves according to

$$x_{t+1} = A \begin{bmatrix} x_t \\ a_t \end{bmatrix} + Bw_{t+1} \tag{2.6}$$

$$w_{t+1} \sim N\left(0, I_{N_w}\right) \tag{2.7}$$

for shocks w_{t+1}, and A has size $N_x \times N_{xa}$ and B has size $N_x \times N_w$.

The private sector has limited information, each period observing N_z signals, collected in z_t,

$$z_t = C \begin{bmatrix} x_t \\ a_t \end{bmatrix} \tag{2.8}$$

$$I_t^{ps} = z^t = \{z_\tau | \tau \leq t\} \tag{2.9}$$

where C has size $N_z \times N_{xa}$.

The policymaker faces a constraint that is, in part, forward looking based on the private sector's information set. At every period t, there are N_μ linear constraints which must hold for every possible realization of shocks,

$$0 = D \begin{bmatrix} x_t \\ a_t \end{bmatrix} + E \left\{ J \begin{bmatrix} x_{t+1} \\ a_{t+1} \end{bmatrix} | I_t^{ps} \right\}$$

where D and J have size $N_\mu \times N_{xa}$. In this paper, I use subscript $|t$ notation to represent based on I_t^{ps}, $x_{t+1|t} = E\left\{x_{t+1}|I_t^{ps}\right\}$.

Definition 1. The *Forward-Looking Constraint* for period t is,

$$0 = D \begin{bmatrix} x_t \\ a_t \end{bmatrix} + J \begin{bmatrix} x_{t+1|t} \\ a_{t+1|t} \end{bmatrix} \tag{2.10}$$

This type of constraint is common in the linearization of equilibrium conditions that determine private sector choices. For instance, in a rational expectations monetary model, the private sector sets prices in the current period based in part about their expectations for prices for the next period.

So that the private sector will use the Kalman update, I require that a_t be mean-zero and jointly normal with the other random variables in the model.[6] This requires that, for a_0, there exists an $N_a \times N_x$ matrix G_0 such that

$$E\left\{a_0|x_0\right\} = G_0 x_0$$

$$\text{Var}\left(E\left\{a_0|x_0\right\}\right) = G_0 \Sigma_0^x G_0^T$$

However, it's possible that a_0 has additional variance beyond the shock-based covariance, i.e. $\Sigma_0^a - G_0 \Sigma_0^x G_0^T \neq 0$. To model extra variance, I introduce a random variable η_0, whose distribution the policymaker chooses. Let Σ_0^η of size $N_a \times N_a$ be defined as

$$\Sigma_0^\eta \equiv \text{Var}\left(a_0\right) - \text{Var}\left(E\left\{a_0|x_0\right\}\right) = \text{Var}\left(a_0\right) - G_0 \Sigma_0^x G_0^T$$

[6] Note that $h_0 = x_0$.

Any a_0 that is jointly normal with x_0 can be expressed as

$$a_0 = G_0 x_0 + \eta_0 \tag{2.11}$$
$$\eta_0 \sim N\left(0, \Sigma_0^\eta\right)$$

with G_0 and Σ_0^η defined as above. The fact that that their could be an η_t shock to a_t means that the full history of shocks is defined as,

$$h_0 \equiv x_0$$
$$\Sigma_0^h = \Sigma_0^x$$
$$h_{t+1} \equiv \begin{bmatrix} h_t \\ \eta_t \\ w_{t+1} \end{bmatrix} \tag{2.12}$$
$$\Sigma_{t+1}^h = \begin{bmatrix} \Sigma_t^h & 0 & 0 \\ 0 & \Sigma_t^\eta & 0 \\ 0 & 0 & I_{N_w} \end{bmatrix}$$

and a full-history dependent a_t can be expressed in the form,

$$a_t \equiv G_t h_t + \eta_t$$
$$\eta_t \sim N\left(0, \Sigma_t^\eta\right)$$

for matrix $N_a \times (N_x + t\,(N_a + N_w))$ matrix G_t and positive, semi-definite $N_a \times N_a$ matrix Σ_t^η. Note that the size of the matrix G_t grows linearly with t.

Instead of looking for state-contingent policies that are jointly normal with the history $\{a_t\,(h_t)\}_{t=0}^\infty$, we equivalently look for a sequence problem of choosing matrices $\{G_t, \Sigma_t^\eta\}_{t=0}^\infty$. This restriction ensures that the private sector can use the Kalman update. These matrices are chosen *before* x_0 or any other shocks realize. At time $t = -1$, the policymaker is choosing the joint covariance of all future a_t with the relevant histories. Note a significant difference from most other models: normally, G_t would be chosen based on information available at time t. That is never the case in this paper. The G_0 that minimizes losses after knowing x_0 is different than the one that minimizes losses knowing only the distribution of x_0, i.e. Σ_0^x. The policymaker is committing to a certain policy and its informational effect on the private sector. *All* the relevant matrices will be chosen based on Σ_0^x or another covariance that is calculable *before* any realizations take place. The underlying model is of stochastic Gaussian shocks and variables, but the sequence problem of choosing matrices is *deterministic*, and based on covariance matrices calculated before any shocks. I call this informational position the *unconditional perspective.*

This contrasts with a full information model, where the policy can be decided after x_0 is realized. The difference is because under full information, a_0 does not give the private sector any additional information. So the optimal $a_0(x_0)$ can be chosen state-by-state. For $\tilde{x}_0 \neq x_0$, the choice of $a_0(x_0)$ does not affect outcomes under $\tilde{x}_0$, and therefore it does not affect the optimal choice $a_0(\tilde{x}_0)$. However, when a_0 informs the private sector, the choice of $a_0(x_0)$ will affect how the private sector interprets z_0 across more than just x_0. Different z_0 updating changes expectations and therefore the optimal $a_0(\tilde{x}_0)$. For these reasons, the optimal policy $a_0(x_0)$ depends on the entire distribution of Σ_0^x, and cannot be calculated in isolation. Fortunately, optimal choices of G_0 and Σ_0^η define the full policy, and are chosen based on Σ_0^x.

To summarize, there are two reasons to prefer the sequence problem of $\{G_t, \Sigma_t^\eta\}_{t=0}^\infty$ to the state-contingent problem of $\{a_t(h_t)\}_{t=0}^\infty$. First, by treating it as a sequence problem of matrices, we guarantee that the private sector will be able to use a Kalman filter to update their beliefs. Second, when a_t informs the private sector, it cannot be solved for state-by-state. That is, we cannot find the optimal mapping $x_0 \to a_0^*$ using only x_0. The entire policy needs to be optimized, and the entire policy for a_0 is captured in the matrices (G_0, Σ_0^η).

Definition 2. A *Full History Sequence* (FHS) is a sequence of matrices, $\{G_t, \Sigma_t^\eta\}_{t=0}^\infty$, where G_t has size $N_a \times (N_x + t(N_a + N_w))$, Σ_t^η is a positive, semi-definite $N_a \times N_a$

matrix. The *set* $\mathcal{FH}$ is the subset of FHS that when used to define a_t,

$$a_t = G_t h_t + \eta_t$$
$$\eta_t \sim N\left(0, \Sigma_t^\eta\right)$$

meets the constraints (2.6), (2.7), (2.8), (2.9), and the Forward-Looking Constraint for all $t \geq 0$.

$$\mathcal{FH} = \{\{G_t, \Sigma_t^\eta\}_{t=0}^\infty : a_t = G_t h_t + \eta_t \implies \text{model constraints are met } \forall t \geq 0\}$$

Definition 3. The Full History Sequence Problem (FHSP) is,

$$V\left(\Sigma_0^x\right) = \min_{\left\{G_t, \Sigma_t^\eta\right\}_{t=0}^\infty \in \mathcal{FH}} E\left\{\sum_{t=0}^\infty \beta^t L_t\right\}$$

for all $t \geq 0$, where Σ_0^x is a positive semi-definite matrix of size $N_x \times N_x$, $\beta \in [0,1)$, L_t is defined in (2.5), x_t evolves according to (2.6), and a_t is defined as in Definition 2.

The FHSP represents a general policy for a_t that is jointly normal with the history of shocks h_t. Solving such a general model is intractable because the number of parameters to choose at time t grows linearly with the length of the history: $N_a\left(N_x + t\left(N_a + N_w\right) + \frac{1}{2}\left(N_a + 1\right)\right)$.[7] Unlike other informational structures or under discretion, the problem cannot be solved by taking a first order condition on a_t. To find the optimal policy we would need to consider the weights G_t and how they affect the information updating of the private sector.

Definition 4. A *Commitment Error Sequence* (CES) is a sequence of matrices, $\{G_t^e, G_t^c, \Sigma_t^\eta\}_{t=0}^\infty$, where G_t^e is a matrix of size $N_a \times N_x$, G_t^c is a matrix of size $N_a \times N_{xa}$,

[7] The final term comes from Σ_t^η being symmetric.

Σ_t^η is a positive, semi-definite $N_a \times N_a$ matrix. The *set* $\mathcal{CE}$ is the domain of CES that when used to define a_t,

$$a_0 = G_0^e x_0 + \eta_0$$

$$a_{t+1} = G_t^c \begin{bmatrix} x_{t|t} \\ a_{t|t} \end{bmatrix} + G_{t+1}^e \left(x_{t+1} - x_{t+1|t} \right) + \eta_{t+1}$$

$$\eta_t \sim N\left(0, \Sigma_t^\eta\right)$$

$$\begin{bmatrix} x_{t+1|t} \\ x_{t|t} \\ a_{t|t} \end{bmatrix} = E\left\{ \begin{bmatrix} x_{t+1} \\ x_t \\ a_t \end{bmatrix} |I_t^{ps} \right\}$$

meets the constraints (2.6), (2.7), (2.8), (2.9), and the Forward-Looking Constraint for all $t \geq 0$.

$$\mathcal{CE} = \{\{G_t^e, G_t^c, \Sigma_t^\eta\}_{t=0}^\infty : a_t \text{ defined above} \implies \text{model constraints are met } \forall t \geq 0\}$$

Definition 5. The Commitment-Error Sequence Problem (CESP) is,

$$W\left(\Sigma_0^x\right) = \min_{\left\{G_t^e, G_t^c, \Sigma_t^\eta\right\}_{t=0}^\infty \in \mathcal{CE}} E\left\{ \sum_{t=0}^{\infty} \beta^t L_t \right\}$$

where Σ_0^x is a positive semi-definite matrix of size $N_x \times N_x$, $\beta \in [0, 1)$, for all $t \geq 0$, L_t is defined in (2.5), x_t evolves according to (2.6), and a_t is defined as in Definition 4.

The CESP has the crucial dimension reduction in the proof. Instead of choosing ever-growing number of weights each period. The policymaker in the CESP is choosing a fixed number of parameters for every t: $N_a \left(2N_x + N_a + \frac{1}{2}\left(N_a + 1\right)\right)$.

Definition 6. An FHS or CES has the *Span Property at time t* iff $a_{t+1|t} \in \text{span}\left(x_{t|t}, a_{t|t}\right)$. That is, there exists a $N_a \times N_{xa}$ matrix G_t^c such that $a_{t+1|t} = G_t^c \begin{bmatrix} x_{t|t} \\ a_{t|t} \end{bmatrix}$.

An FHS or CES has the *Span Property* iff it has the *Span Property* for all $t \geq 0$.

By construction, all CES have the Span Property. In words, it requires that the predictable behavior of a_{t+1}, given the private sector's information at t, to depend only on the private sector's estimate of x_t and a_t. By taking expectations and rearranging, the Forward-Looking Constraint for time t can be reformulated as two separate constraints

$$J = \begin{bmatrix} J_x & J_a \end{bmatrix}$$

$$0 = (D + J_x A) \begin{bmatrix} x_{t|t} \\ a_{t|t} \end{bmatrix} + J_a a_{t+1|t}$$

$$0 = D \begin{bmatrix} x_t - x_{t|t} \\ a_t - a_{t|t} \end{bmatrix}$$

The formulation above shows the intuition for why an optimal FHS would have the Span Property. Extra variance in a_{t+1} is weakly costly, and therefore so is extra variation in $a_{t+1|t}$. However, it is useful for $a_{t+1|t}$ to respond to $x_{t|t}$ and $a_{t|t}$. Proposition 13 shows that any FHS without the Span Property at time t can be weakly improved by modifying the sequence after $t+1$ so that it does have the Span Property at time t. Proposition 16 shows that any FHS with the Span Property can have its expected losses $\{E\{L_t\}\}_{t=0}^{\infty}$ matched at each period by a CES.

Recursive Formulation of Constraints

Above, all constraints were phrased in terms of specific periods 0, $t-1$, t, or $t+1$. The VBR will be recursive, so in this subsection I present the equivalent recursive version of the constraints. In this subsection, subscript p will indicate the previous period, c the current, and n the next. Policy will be determined based on augmented state y_c, defined below, and the matrices committed to based on the unconditional distribution of y_c, before any shocks are realized. All of the constraints and intermediate values can be expressed in terms of y_c, and the shocks before y_n: η_c and w_n. That alternative form requires a fair bit of linear algebra book keeping, and it is done in Appendix A.1.

The augmented state y_c has dimension $N_y \equiv 2N_x + N_a$ and is defined as,

$$y_c \equiv \begin{bmatrix} x_c \\ x_{c|p} \\ a_{c|p} \end{bmatrix} \sim N\left(0, \Sigma^y\right)$$

and has the distribution implied by the covariance of those three random variables, with

$$\begin{bmatrix} x_{c|p} \\ a_{c|p} \end{bmatrix} = E\left\{ \begin{bmatrix} x_c \\ a_c \end{bmatrix} |I_p^{ps} \right\}$$

Augmented state y_c is composed of the current backward-looking state x_c, the previous period's private-sector expectation of x_c, and the previous period's private-sector expectation of the current action a_c. The final element, $a_{c|p}$, is the commitment from the previous period that must be followed through upon, i.e. a_c must be chosen so that the $a_{c|p}$ from the previous period was rational.

The action for period c will be determined as follows,

$$a_c = a_{c|p} + G^e\left(\Sigma^y\right)\left(x_c - x_{c|p}\right) + \eta_c \tag{2.13}$$
$$\eta_c \sim N\left(0, \Sigma^\eta\left(\Sigma^y\right)\right)$$

where $G^e\left(\Sigma^y\right)$ is size $N_a \times N_x$, and $\Sigma^\eta\left(\Sigma^y\right)$ is positive semi-definite size $N_a \times N_a$, and both were chosen by the policymaker from the unconditional perspective based on the distribution of y_c, Σ^y. Note that, for all y_c, $E\left\{x_c - x_{c|p}|I_p^{ps}\right\} = 0$ and $E\left\{\eta_c|I_p^{ps}\right\} = 0$. Therefore, by constructing a_c in this fashion, it guarantees that $E\left\{a_c|I_p^{ps}\right\} = \begin{bmatrix} 0 & 0 & I_{N_a} \end{bmatrix} y_c$. One aspect of the proof is that the distribution Σ^y is sufficient for the policymaker to choose the an optimal G^e and Σ^η that performs as well as a FHS.

All of the numbered equations in this section can be calculated explicitly in terms of y_c, Σ^y, η_c, Σ^η, G^e, and G^c. I perform these calculations in Appendix A.1.

Here are the recursive versions of the constraints that the functions must meet,

$$L_c = \begin{bmatrix} x_c \\ a_c \end{bmatrix}^T L \begin{bmatrix} x_c \\ a_c \end{bmatrix} \tag{2.14}$$

$$x_n = A \begin{bmatrix} x_c \\ a_c \end{bmatrix} + Bw_n \tag{2.15}$$

$$w_n \sim N\left(0, I_{N_w}\right) \tag{2.16}$$

The private sector's information is no longer the history of shocks. Instead it is the prior period's estimates, that are part of y_c, and the observed signal today,

$$z_c = C \begin{bmatrix} x_c \\ a_c \end{bmatrix} \tag{2.17}$$

$$I_c^{ps} = \left\{x_{c|p}, a_{c|p}, z_c\right\} \tag{2.18}$$

The private sector will use a Kalman filter to update their beliefs

$$\begin{aligned} \begin{bmatrix} x_{c|c} \\ a_{c|c} \end{bmatrix} &= E\left\{ \begin{bmatrix} x_c \\ a_c \end{bmatrix} |I_c^{ps} \right\} \\ &= \begin{bmatrix} x_{c|p} \\ a_{c|p} \end{bmatrix} + K^{xa}\left(z - C \begin{bmatrix} x_{c|p} \\ a_{c|p} \end{bmatrix}\right) \end{aligned} \tag{2.19}$$

where the K^{xa} depends on Σ^y, G^e, and Σ^η, and its formula is given in equation (A.2) in Appendix A.1.

For input to the recursive call, I require

$$\begin{bmatrix} x_{n|c} \\ a_{n|c} \end{bmatrix} = \begin{bmatrix} A \\ G^c\left(\Sigma^y\right) \end{bmatrix} \begin{bmatrix} x_{c|c} \\ a_{c|c} \end{bmatrix} \tag{2.20}$$

where $G^c\left(\Sigma^y\right)$ size $N_a \times N_{xa}$ and is the third matrix chosen by the policymaker from the unconditional perspective. Together, equations (2.15) and (2.20) define y_n, and therefore its distribution as well.

$$y_n \equiv \begin{bmatrix} x_n \\ x_{n|c} \\ a_{n|c} \end{bmatrix} \tag{2.21}$$

$$y_n \sim N\left(0, \Sigma_n^y\right)$$

with equation (A.4) in Appendix A.1 having an explicit definition of y_n and Σ^y_n.

This is consistent with how $a_n\left(\Sigma^y_n\right)$ will be constructed (changing subscript to be consistent with this period from c to n and p to c):

$$a_n\left(\Sigma^y_n\right) = a_{n|c} + G^e\left(\Sigma^y_n\right)\left(x_n - x_{n|c}\right) + \eta_n$$

$$E\left\{a_n\left(\Sigma^y_n\right)|I^{ps}_c\right\} = a_{n|c}$$

For all potential Σ^y, it must be the case that the recursive version of the Forward-Looking Constraint also holds,

$$\begin{aligned} 0 &= D\begin{bmatrix} x_c \\ a_c \end{bmatrix} + J\begin{bmatrix} x_{n|c} \\ a_{n|c} \end{bmatrix} \\ &= D\begin{bmatrix} x_c \\ a_c \end{bmatrix} + J\begin{bmatrix} A \\ G^c \end{bmatrix}\begin{bmatrix} x_{c|c} \\ a_{c|c} \end{bmatrix} \end{aligned} \tag{2.22}$$

which the appendix shows how to check for a given Σ^y. In particular, Lemma 24 proves that from Σ^y, G^e, G^c, and Σ^η we can check whether the Forward-Looking Constraint holds for all possible $\begin{bmatrix} y_c \\ \eta_c \end{bmatrix}$.

Definition 7. Define $\mathcal{VB}$ as the domain of valid functions for $G^e\left(\Sigma^y\right), G^c\left(\Sigma^y\right), \Sigma^\eta\left(\Sigma^y\right)$ for the Value Bellman Problem, in the next definition. They are used to define $a_c\left(\Sigma^y\right)$. Let Σ^y be such that

$$y_c = \begin{bmatrix} x_c \\ x_{c|p} \\ a_{c|p} \end{bmatrix} \sim N\left(0, \Sigma^y\right)$$

and define a_c as

$$a_c = a_{c|p} + G^e\left(\Sigma^y\right)\left(x_c - x_{c|p}\right) + \eta_c$$

$$\eta_c \sim N\left(0, \Sigma^\eta\left(\Sigma^y\right)\right)$$

where Σ^y is a positive, semi-definite $N_y \times N_y$ matrix, $G^e\left(\Sigma^y\right)$ has size $N_a \times N_x$, $G^c\left(\Sigma^y\right)$ has size $N_a \times N_{xa}$, $\Sigma^\eta\left(\Sigma^y\right)$ is a positive, semi-definite matrix of size $N_a \times N_a$.

Constraints (2.14), (2.16), (2.17), (2.18), and (2.22) are met for all possible $\begin{bmatrix} y_c \\ \eta_c \end{bmatrix}$.

$$\mathcal{VB} = \Big\{ G^e\left(\Sigma^y\right), G^c\left(\Sigma^y\right), \Sigma^\eta\left(\Sigma^y\right) :$$

$$a_c \text{ as defined above} \implies \text{model constraints are met } \forall \Sigma^y, \begin{bmatrix} y_c \\ \eta_c \end{bmatrix} \Big\}$$

Definition 8. The Variance Bellman Problem (VBP) is,

$$U\left(\Sigma^y\right) = \min_{\{G^e, G^c, \Sigma^\eta\} \in \mathcal{VB}} E\left\{L_c\right\} + \beta U\left(\Sigma^y_n\right)$$

$$y_n \sim N\left(0, \Sigma^y_n\right)$$

where Σ^y is a positive, semi-definite matrix of size $(2N_x + N_a) \times (2N_x + N_a)$, $\beta \in [0, 1)$, L_c as defined in (2.14), the definition of $\mathcal{VB}$ ensures all constraints are met, and y_n is defined in (2.21).

The VBP has policy functions have the same dimensions as one period of a CES, with $N_a\left(2N_x + N_a + \frac{1}{2}\left(N_a + 1\right)\right)$ weights. The key difference is that they are functions instead of sequences. The domain of the functions is possible augmented state covariance matrices for any period, Σ^y. In the VBP formulation, the choices of matrices this period trade off costs in terms of losses for this period with the discounted losses of all future periods, represented by $U\left(\Sigma^y_n\right)$. Normal recursive effects of G^e and Σ^η are that they will affect costs today and the transition and therefore distribution for x_n. The additional effect due to signaling is that they will affect the information in z_c and therefore the accuracy of $x_{n|c}$, which is part of the next augmented state y_n. Via commitment, G^c determines $a_{n|c}$ which can mitigate this period's Forward-Looking Constraint, and is also part of y_n with associated future costs.

Note U is recursive in the covariances through time. It represents the discounted expected losses for the infinite problem. It formed by the expected loss for this period plus the discounted infinite losses starting next period. At its root, however is choosing the optimal matrices for the specific period given augmented state covariance matrix Σ^y for that period. Like the FHSP and CESP, all the matrices are chosen from the unconditional perspective, *before* any shocks realize. The VBP is a convenient formulation for choosing the optimal matrices recursively, but those matrices are still chosen at $t = -1$.

2.4 Equivalence of FHSP, CESP, and VBP

In this paper I assume existence of optimal solutions to FHSP, CESP, and VBP, i.e. there exists at least one minimand for each problem and the value functions are well defined. Below, I outline the way that the problems are proven to be equivalent with

$$V\left(\Sigma_0^x\right) = W\left(\Sigma_0^x\right) = U\left(\begin{bmatrix} \Sigma_0^x & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\right)$$

The equivalence of FHSP and CESP is shown in Theorem 18. It uses corollary 14 to show that there exists an optimal FHS with the Span Property. Then, Proposition 16 shows that any FHS with the Span Property can have its expected losses matched with a properly constructed CES. Combined these results show that the optimal FHS cannot outperform the optimal CES. Proposition 17 shows that any CES has a corresponding FHS that has the same a_t and therefore losses. Thus, the optimal CES cannot outperform the optimal FHS, and the problems are equivalent.

The equivalence of CESP and VBP is shown in Theorem 19, which is simpler. It does this by showing that expectations work appropriately and that Σ^y is a sufficient statistic for the recursive part of VBP.

The most complicated part of the proof is the one on which corollary 14 is based: Proposition 13 showing that an FHS in $\mathcal{FH}$ without the Span Property can be weakly improved into one that has the Span Property. Proposition 13 relies on changing the behavior on the subspace that violates the Span Property.

2.4.1 Proof of Equivalence

Lemma 9. *Let FHS* $\{G_t, \Sigma_t^\eta\}_{t=0}^\infty \in \mathcal{FH}$*. The private-sector's estimation of the history depends on the matrices up to* t*, and forms a projection matrix* $P_{t+1|t}\left(\{G_\tau, \Sigma_\tau^\eta\}_{\tau=0}^t\right)$ *of size* $N_{h,t+1} \times N_{h,t+1}$*,*

$$
\begin{aligned}
E\left\{h_{t+1}|I_t^{ps}\right\} &= h_{t+1|t} \\
&= P_{t+1|t}h_{t+1} \\
h_{t+1|t} &= P_{t+1|t}h_{t+1|t}
\end{aligned}
$$

Proof. First, construct the $N_x \times N_{ht}$ matrix φ_t inductively, so that $x_t = \varphi_t h_t$. $h_0 = x_0$, so $\varphi_0 = I_{N_x}$.

$$
\begin{aligned}
x_{t+1} &= A\begin{bmatrix} x_t \\ a_t \end{bmatrix} + Bw_{t+1} \\
&= A\left(\begin{bmatrix} \varphi_t \\ G_t \end{bmatrix} h_t + \begin{bmatrix} 0 \\ I_{N_a} \end{bmatrix} \eta_t\right) + Bw_{t+1} \\
h_{t+1} &= \begin{bmatrix} h_t \\ \eta_t \\ w_{t+1} \end{bmatrix} \\
\varphi_{t+1} &= \begin{bmatrix} A\begin{bmatrix} \varphi_t & 0 \\ G_t & I_{N_a} \end{bmatrix} & B \end{bmatrix}
\end{aligned}
$$

Initially, $h_{0|-1} = x_{0|-1} = 0$, so $P_{0|-1} = 0_{N_x \times N_x}$. Now the inductive step is to use $P_{t|t-1}$, φ_t, G_t, and Σ_t^η to construct $P_{t+1|t}$.

First calculate how $h_{t+1|t-1}$ depends on the prior $P_{t|t-1}$.

$$h_{t+1|t-1} = \begin{bmatrix} h_{t|t-1} \\ 0 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} P_{t|t-1} \\ 0 \\ 0 \end{bmatrix} h_t$$

$$= P_{t+1|t-1} h_{t+1}$$

$$P_{t+1|t-1} = \begin{bmatrix} P_{t|t-1} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

where $P_{t+1|t-1}$ has size $N_{h,t+1} \times N_{h,t+1}$ because $w_{t+1|t-1} = 0$ and $\eta_{t|t-1} = 0$.

Now consider the information effect of z_t.

$$z_t = C \begin{bmatrix} x_t \\ a_t \end{bmatrix}$$

$$= C_t h_{t+1}$$

$$C_t \equiv C \begin{bmatrix} \varphi_t & 0 & 0 \\ G_t & I_{N_a} & 0 \end{bmatrix}$$

where C_t has size $N_z \times N_{h,t+1}$, and the last column of the final matrix is 0s because w_{t+1} does not affect z_t (but η_t does, hence h_{t+1} instead of h_t).

Now we can calculate $z_{t|t-1}$, and the prediction error,

$$z_{t|t-1} = C_t h_{t+1|t-1} = C_t P_{t+1|t-1} h_{t+1}$$

$$z_t - z_{t|t-1} = C_t \left(h_{t+1} - h_{t+1|t-1} \right)$$

In this paper, when calculating uncertainties with respect to the private sector's information for any variable b, I will use the form Σ_t^b to represent the variance of b_t, $\Sigma_{t|t-1}^b$ to represent the variance of the private sector's estimate $b_{t|t-1}$, and $\Sigma_{t|t-1}^{b|ps}$ to represent the variance of the private sector's prediction error $\left(b_t - b_{t|t-1}\right)$. Applying that notation here,

$$\Sigma^h_{t+1} = \begin{bmatrix} \Sigma^h_t & 0 & 0 \\ 0 & \Sigma^\eta_t & 0 \\ 0 & 0 & \Sigma^w \end{bmatrix}$$

$$\text{Var}\left(h_{t+1|t-1}\right) = \Sigma^h_{t+1|t-1} = P_{t+1|t-1}\Sigma^h_{t+1}P^T_{t+1|t-1}$$

$$\text{Var}\left(h_{t+1} - h_{t+1|t-1}\right) = \Sigma^{h|ps}_{t+1|t-1} = \Sigma^h_{t+1} - \Sigma^h_{t+1|t-1}$$

$$\text{Var}\left(z_t - z_{t|t-1}\right) = \Sigma^{z|ps}_{t|t-1} = C_t\Sigma^{h|ps}_{t+1|t-1}C^T_t$$

where Σ^h_{t+1}, $\Sigma^h_{t+1|t-1}$, and $\Sigma^{h|ps}_{t+1|t-1}$ are positive, semi-definite matrices of size $N_{h,t+1} \times N_{h,t+1}$; $\Sigma^{z|ps}_{t|t-1}$ is a positive, semi-definite matrix of size $N_z \times N_z$.

The Kalman update for the private sector will have the form

$$h_{t+1|t} = h_{t+1|t-1} + K^h_t\left(z_t - z_{t|t-1}\right)$$

where K^h_t has size $N_{h,t+1} \times N_z$. I show in Appendix A.2 how to calculate an optimal K^h_t even if $\Sigma^{z|ps}_{t|t-1}$ is not full rank,

$$K^h_t \equiv \Sigma^{h|ps}_{t+1|t-1}C^T_t\left(\Sigma^{z|ps}_{t|t-1}\right)^+$$

where $\left(\Sigma^{z|ps}_{t|t-1}\right)^+$ is the Moore–Penrose pseudoinverse. The prediction then becomes

$$\begin{aligned} h_{t+1|t} &= h_{t+1|t-1} + K^h_t\left(C_t h_{t+1} - C_t h_{t+1|t-1}\right) \\ &= K^h_t C_t h_{t+1} + \left(I - K^h_t C_t\right)h_{t+1|t-1} \\ &= P_{t+1|t}h_{t+1} \\ P_{t+1|t} &= \left(K^h_t C_t + \left(I - K^h_t C_t\right)P_{t+1|t-1}\right) \end{aligned}$$

Through the series of calculations, $P_{t+1|t}$ depended on G_t, Σ^η_t, φ_t, and $P_{t|t-1}$. Thus, it is a function of $\{G_\tau, \Sigma^\eta_\tau\}^t_{\tau=0}$. □

The prior lemma was a matter of accumulating policies G_t and Σ_t^η, to track how the private sector would calculate their Kalman updates based on z_t. From that, along with the dynamics of x_t, we can calculate the projection $P_{t+1|t}$ that maps shocks to the private sector's estimates of those shocks.

In the next proposition, I will be using a fact about unconditional expectations of mean-zero Gaussian linear systems.

Lemma 10. *In general, the losses at time t will be*

$$E\left\{L_t\right\} = E\left\{\begin{bmatrix}x_t\\a_t\end{bmatrix}^T L\begin{bmatrix}x_t\\a_t\end{bmatrix}\right\} = \left\langle L,\, Var\left(\begin{bmatrix}x_t\\a_t\end{bmatrix}\right)\right\rangle_F$$

where $\langle A, B\rangle_F$ is the Frobenius inner product, defined as $\langle A, B\rangle_F = \sum_{i,j} a_{ij}b_{ij}$.

Proof. Let b be a mean-zero random vector of length N_b, c be a mean-zero random vector of length N_c, and M be a constant matrix of size $N_b \times N_c$,

$$\begin{aligned}E\left\{b^T Mc\right\} &= E\left\{\sum_{i,j} b_i M_{ij} c_j\right\}\\ &= \sum_{i,j} M_{ij} E\left\{b_i c_j\right\}\\ &= \langle M, \operatorname{Cov}(b,c)\rangle_F\end{aligned}$$

□

Definition 11. Let $\tau \geq 0$, and $t \geq \tau + 1$. Let H_t be the vector space of potential histories h_t. $h_{t|\tau}$, x_τ, and a_τ are linear mappings on that space

$$h_{t|\tau} = \begin{bmatrix}h_{\tau+1|\tau}\\0\end{bmatrix} = \begin{bmatrix}P_{\tau+1|\tau} & 0\\0 & 0\end{bmatrix} h_t$$

$$x_\tau = \varphi_\tau h_\tau = \begin{bmatrix}\varphi_\tau & 0\end{bmatrix} h_t$$

$$a_\tau = G_\tau h_\tau + \eta_\tau = \begin{bmatrix}G_\tau & I_{N_a} & 0\end{bmatrix} h_t$$

Define the subspace

$$H_t^{*\tau} \equiv \left\{ h_t : h_{t|\tau} = h_t \wedge x_\tau = 0 \wedge a_\tau = 0 \right\}$$

The superscript τ indicates the constraints depend on x, a, and I^{ps} from period τ. Because all the constraints are based on linear conditions, $H_t^{*\tau}$ is a subspace and has an associated projection $P_t^{*\tau}$. H_t can be partitioned into $H_t^{*\tau}$ and its orthogonal compliment, $H_t^{*\tau\perp}$,

$$h_t = h_t^{*\tau} + h_t^{*\tau\perp}$$

$$h_t^{*\tau} = P_t^{*\tau} h_t \in H_t^{*\tau}$$

$$h_t^{*\tau\perp} = \left(I - P_t^{*\tau}\right) h_t \in H_t^{*\tau\perp}$$

Likewise, any vector x_{t-1} or a_{t-1}[8] can be decomposed into its projection onto $H_t^{*\tau}$ and $H_t^{*\tau\perp}$:

$$\begin{aligned} a_{t-1} &= \begin{bmatrix} G_{t-1} & I_{N_a} & 0 \end{bmatrix} h_t \\ &= a_{t-1}^{*\tau} + a_{t-1}^{*\tau\perp} \\ a_{t-1}^{*\tau} &= \begin{bmatrix} G_{t-1} & I_{N_a} & 0 \end{bmatrix} h_t^{*\tau} \\ &= \begin{bmatrix} G_{t-1} & I_{N_a} & 0 \end{bmatrix} P_t^{*\tau} h_t \\ a_{t-1}^{*\tau\perp} &= \begin{bmatrix} G_{t-1} & I_{N_a} & 0 \end{bmatrix} \left(I - P_t^{*\tau}\right) h_t \end{aligned}$$

and similarly for $x_{t-1} = \begin{bmatrix} \varphi_{t-1} & 0 & 0 \end{bmatrix} h_t$. □

There are two special features of $H_{\tau+1}^{*\tau}$. The first is that expectations based on I_τ^{ps} are entirely accurate. It is a subspace of the space of the expectation projection from Lemma 9,

$$h_{\tau+1}^{*\tau} \in H_{\tau+1}^{*\tau} \implies h_{\tau+1}^{*\tau} = P_{\tau+1|\tau} h_{\tau+1}^{*\tau}$$

This means that $P_{\tau+1}^{*\tau} = P_{\tau+1}^{*\tau} P_{\tau+1|\tau} = P_{\tau+1|\tau} P_{\tau+1}^{*\tau}$.

[8] H_t is necessary to specify a_{t-1}, because a_{t-1} depends in part on η_{t-1} which is not in h_{t-1}.

Second, note also that for all $t \geq \tau+1$, $h_t^{*\tau} \in H_t^{*\tau}$ implies that for all s such that $t \geq s \geq \tau+1$ inside $h_t^{*\tau}$, $w_s = 0$, and for all $t > s' \geq \tau+1$ $\eta_{s'} = 0$. (η_t is not inside h_t).

Lemma 12. *An FHS* $\{G_t, \Sigma_t^\eta\}_{t=0}^\infty$ *has the Span Property at time* $\tau \iff a_{\tau+1|\tau}^{*\tau} = 0$ *for all* $h_{\tau+1} \in H_{\tau+1}$.

Proof. $\implies$ Let $h_{\tau+1} \in H_{\tau+1}$, and let $h_{\tau+1}^{*\tau} = P_{\tau+1}^{*\tau} h_{\tau+1}$. One of the conditions of $H_{\tau+1}^{*\tau}$ is that predictions based on I_τ^{ps} are accurate, . Therefore, $h_{\tau+1}^{*\tau} = P_{\tau+1|\tau} h_{\tau+1}^{*\tau}$. Combining with the fact that within $H_{\tau+1}^{*\tau}$, $x_\tau = a_\tau = 0$,

$$\begin{aligned} x_{\tau|\tau}^{*\tau} &= \begin{bmatrix} \varphi_t & 0 & 0 \end{bmatrix} P_{\tau+1|\tau} h_{\tau+1}^{*\tau} \\ &= \begin{bmatrix} \varphi_t & 0 & 0 \end{bmatrix} h_{\tau+1}^{*\tau} = x_\tau^{*\tau} = 0 \\ a_{\tau|\tau}^{*\tau} &= \begin{bmatrix} G_\tau & I_{N_a} & 0 \end{bmatrix} P_{\tau+1|\tau} h_{\tau+1}^{*\tau} \\ &= \begin{bmatrix} G_\tau & I_{N_a} & 0 \end{bmatrix} h_{\tau+1}^{*\tau} = a_\tau^{*\tau} = 0 \end{aligned}$$

thus, $\operatorname{span}\left(x_{\tau|\tau}^{*\tau}, a_{\tau|\tau}^{*\tau}\right) = \operatorname{span}(\{0\})$, and the Span Property implies $a_{\tau+1|\tau}^{*\tau} = 0$.

$\impliedby$ By proving the contrapositive $\neg A \implies \neg B$. Assume the FHS does not have the Span Property at time τ. Define $v_{\tau+1|\tau} \equiv a_{\tau+1|\tau} - E\left\{a_{\tau+1|\tau} | x_{\tau|\tau}, a_{\tau|\tau}\right\}$. Note, by construction $E\left\{v_{\tau+1|\tau} | x_{\tau|\tau}, a_{\tau|\tau}\right\} = 0$, i.e. $v_{\tau+1|\tau}$ is independent of $\left(x_{\tau|\tau}, a_{\tau|\tau}\right)$.

By the assumption, it must be the case that $\operatorname{Var}\left(v_{\tau+1|\tau}\right) \neq 0$. Let P_v be the $N_a \times N_{h,\tau+1}$ mapping from $h_{\tau+1|\tau}$ to $v_{\tau+1|\tau}$, so $v_{\tau+1|\tau} = P_v P_{\tau+1|\tau} h_{\tau+1}$. Because $v_{\tau+1|\tau}$ is independent of $\left(x_{\tau|\tau}, a_{\tau|\tau}\right)$ and has non-zero variance, there must exist an $h_{\tau+1}$ such that 3 things are true

$$\begin{aligned} v_{\tau+1|\tau} &= P_v P_{\tau+1|\tau} h_{\tau+1} \neq 0 \\ a_{\tau|\tau} &= \begin{bmatrix} G_\tau & I_{N_a} & 0 \end{bmatrix} P_{\tau+1|\tau} h_{\tau+1} = 0 \\ x_{\tau|\tau} &= \begin{bmatrix} \varphi_t & 0 & 0 \end{bmatrix} P_{\tau+1|\tau} h_{\tau+1} = 0 \end{aligned}$$

Let $h_{\tau+1|\tau} \equiv P_{\tau+1|\tau} h_{\tau+1}$. By the conditions on $h_{\tau+1}$, three things are true:

$$h_{\tau+1|\tau} = P_{\tau+1|\tau} h_{\tau+1|\tau}$$
$$0 = \begin{bmatrix} G_\tau & I_{N_a} & 0 \end{bmatrix} h_{\tau+1|\tau}$$
$$0 = \begin{bmatrix} \varphi_t & 0 & 0 \end{bmatrix} h_{\tau+1|\tau}$$

Combined, they imply $h_{\tau+1|\tau} \in H^{*\tau}_{\tau+1}$. The $a^{*\tau}_{\tau+1|\tau}$ for $h_{\tau+1|\tau}$ is

$$\begin{aligned} a_{\tau+1|\tau} &= G_{\tau+1} h_{\tau+1|\tau} \\ &= v_{\tau+1|\tau} \neq 0 \end{aligned}$$

□

This lemma shows how $H^{*\tau}_{\tau+1}$ is a useful subspace of $H_{\tau+1}$. Is is precisely the one on which violations of the Span Property for time τ can take place. The key step in the next proposition is to change $a_{\tau+1}$ to be zero on that subspace so that the constructed FHS has the Span Property at time τ.

Proposition 13. *Let FHS $\{G_t, \Sigma^\eta_t\}_{t=0}^\infty \in \mathcal{FH}$ such that it does not have the Span Property at time $\tau \geq 0$. Then, there exists a new FHS $\left\{\tilde{G}_t, \Sigma^\eta_t\right\}_{t=0}^\infty \in \mathcal{FH}$ with the Span Property at τ and weakly lower costs.*

Proof. This is a constructive proof, as we will build the new sequence using the projections onto $H^{*\tau}_t$, for $t \geq \tau + 1$.

The original FHS had

$$a_{\tau+1} = G_{\tau+1} h_{\tau+1} + \eta_{\tau+1}$$

and

$$a_{\tau+1|\tau} = G_{\tau+1} h_{\tau+1|\tau} = G_{\tau+1} P_{\tau+1|\tau} h_{\tau+1}$$

Because the original FHS does not have the Span Property at τ, the lemma above shows that when decomposed

$$\exists h^{*\tau}_{\tau+1} \in H^{*\tau}_{\tau+1} \text{ such that } a^{*\tau}_{\tau+1|\tau} = G_{\tau+1} h^{*\tau}_{\tau+1} \neq 0$$

For all time, the new FHS will have the same $\{\Sigma_t^\eta\}_{t=0}^\infty$. Additionally, the new FHS will have the same choices for $\forall t \leq \tau$, $\tilde{G}_t = G_t$. Note, the full set of subspaces $\{H_t^{*\tau}\}_{t=\tau+1}^\infty$ depend on I_τ^{ps}. I_τ^{ps} hasn't changed, because as Lemma 9 shows, $P_{\tau+1|\tau}$ depends only on $\{G_t, \Sigma_t^\eta\}_{t=0}^\tau$, and those matrices remain the in the new FHS.

The new FHS for time $\tau + 1$ is

$$\tilde{a}_{\tau+1} = \tilde{G}_{\tau+1} h_{\tau+1} + \eta_{\tau+1}$$
$$\tilde{G}_{\tau+1} \equiv G_{\tau+1} \left(I - P_{\tau+1}^{*\tau}\right)$$

To see how the new $\tilde{G}_{\tau+1}$ affects things, decompose $\tilde{a}_{\tau+1|\tau}$ as discussed in Definition 11, starting first with $\tilde{a}_{\tau+1|\tau}^{*\tau}$,

$$\begin{aligned}
\tilde{a}_{\tau+1|\tau} &= \tilde{a}_{\tau+1|\tau}^{*\tau} + \tilde{a}_{\tau+1|\tau}^{*\tau\perp} \\
\tilde{a}_{\tau+1|\tau}^{*\tau} &= \tilde{G}_{\tau+1} P_{\tau+1|\tau} P_{\tau+1}^{*\tau} h_{\tau+1} \\
&= \tilde{G}_{\tau+1} P_{\tau+1}^{*\tau} h_{\tau+1} \\
&= G_{\tau+1} \left(I - P_{\tau+1}^{*\tau}\right) P_{\tau+1}^{*\tau} h_{\tau+1} \\
&= G_{\tau+1} 0 = 0
\end{aligned}$$

with the third equation being because $P_{\tau+1|\tau} P_{\tau+1}^{*\tau} = P_{\tau+1}^{*\tau}$. By right multiplying $G_{\tau+1}$ by $\left(I - P_{\tau+1}^{*\tau}\right)$, $\tilde{a}_{\tau+1|\tau}^{*\tau} = 0$. By the lemma, this means that the new FHS $\left\{\tilde{G}_t\right\}_{t=0}^\infty$ will have the Span Property at time τ.

See that $\tilde{a}_{\tau+1|\tau}^{*\tau} = 0$ complies with the Forward-Looking Constraint at time τ because, $x_\tau^{*\tau} = 0$, $a_\tau^{*\tau} = 0$, $x_{\tau|\tau}^{*\tau} = 0$, $a_{\tau|\tau}^{*\tau} = 0$, $x_{\tau+1|\tau}^{*\tau} = A \begin{bmatrix} x_{\tau|\tau}^{*\tau} \\ a_{\tau|\tau}^{*\tau} \end{bmatrix} = 0$,

$$0 = D \begin{bmatrix} x_\tau^{*\tau} \\ a_\tau^{*\tau} \end{bmatrix} + J \begin{bmatrix} x_{\tau+1|\tau}^{*\tau} \\ \tilde{a}_{\tau+1|\tau}^{*\tau} \end{bmatrix}$$

$H_{\tau+1}^{*\tau}$ is constructed so that all the terms in the constraint besides $\tilde{a}_{\tau+1|\tau}^{*\tau}$ are 0.

Now consider what happens with $\tilde{a}^{*\tau\perp}_{\tau+1|\tau}$. First, observe that $P^{*\tau}_{\tau+1}P_{\tau+1|\tau} = P^{*\tau}_{\tau+1}$. Therefore,

$$\begin{aligned}
\tilde{a}^{*\tau\perp}_{\tau+1|\tau} &= \tilde{G}_{\tau+1}P_{\tau+1|\tau}\left(I - P^{*\tau}_{\tau+1}\right)h_{\tau+1} \\
&= G_{\tau+1}\left(I - P^{*\tau}_{\tau+1}\right)P_{\tau+1|\tau}\left(I - P^{*\tau}_{\tau+1}\right)h_{\tau+1} \\
&= G_{\tau+1}\left(P_{\tau+1|\tau} - P^{*\tau}_{\tau+1}\right)\left(I - P^{*\tau}_{\tau+1}\right)h_{\tau+1} \\
&= G_{\tau+1}P_{\tau+1|\tau}\left(I - P^{*\tau}_{\tau+1}\right)h_{\tau+1} \\
&= G_{\tau+1}P_{\tau+1|\tau}h^{*\tau\perp}_{\tau+1} \\
&= a^{*\tau\perp}_{\tau+1|\tau}
\end{aligned}$$

In other words, for $h^{*\tau\perp}_{\tau+1}$ in the orthogonal compliment, the expected behavior is exactly the same as before. Thus the Forward-Looking Constraint also holds on $H^{*\tau\perp}_{\tau+1}$, because the $\tilde{a}^{*\tau\perp}_{t+1}$ behavior is unchanged, $\tilde{a}^{*\tau\perp}_{\tau+1|\tau} = a^{*\tau\perp}_{\tau+1|\tau}$

$$D\begin{bmatrix} x^{*\tau\perp}_{\tau} \\ a^{*\tau\perp}_{\tau} \end{bmatrix} + J\begin{bmatrix} x^{*\tau\perp}_{\tau+1|\tau} \\ \tilde{a}^{*\tau\perp}_{\tau+1|\tau} \end{bmatrix} = D\begin{bmatrix} x^{*\tau\perp}_{\tau} \\ a^{*\tau\perp}_{\tau} \end{bmatrix} + J\begin{bmatrix} x^{*\tau\perp}_{\tau+1|\tau} \\ a^{*\tau\perp}_{\tau+1|\tau} \end{bmatrix} = 0$$

To summarize, we've shown so far that we generate $\left\{\tilde{G}_t\right\}_{t=0}^{\tau+1}$ through $\tau+1$

$$\tilde{G}_t = \begin{cases} G_t & t \leq \tau \\ G_t\left(I - P^{\tau*}_{\tau+1}\right) & t = \tau+1 \end{cases}$$

and that this FHS has the Span Property at time τ, as well as meeting the τ-period Forward-Looking Constraint. We also showed that when partitioned into the subspace $H^{*\tau}_{\tau+1}$ and its orthogonal compliment

$$x^{*\tau}_{\tau} = x^{*\tau}_{\tau|\tau} = 0$$

$$a^{*\tau}_{\tau} = a^{*\tau}_{\tau|\tau} = 0$$

$$x^{*\tau}_{\tau+1} = x^{*\tau}_{\tau+1|\tau} = 0$$

By Lemma 12, because $G_{\tau+1}$ did not have the Span Property, there $\exists h_{\tau+1}.a^*_{\tau+1|\tau}(h_{\tau+1}) \neq 0$. Which is not the case for $\tilde{G}_{\tau+1}$. When constructing the rest of the FHS, $\left\{\tilde{G}_t\right\}_{t=\tau+2}^{\infty}$, we must deal with the fact that

$$a^{*\tau}_{\tau+1} = G_{\tau+1}h^{*\tau}_{\tau+1} \neq 0$$

could have follow on consequences for subsequent $x_{\tau+2}$, $a_{\tau+2}$, $x_{\tau+3}$, etc. We construct $\left\{\tilde{G}_t\right\}_{t=\tau+2}^{\infty}$ that are consistent with the new $\tilde{G}_{\tau+1}$.

One of the conditions of $H^{*\tau}_t$ is that they have private-sector perfect foresight, i.e. expectations based on I^{ps}_τ are perfectly accurate. This implies $w_{t|\tau} = 0$ and $\eta_{t|\tau} = 0$ for all $t \geq \tau + 1$. We use the corresponding projections $P^{*\tau}_t$ to define all following $\tilde{G}_t$ as we did $\tilde{G}_{\tau+1}$,

$$\tilde{G}_t \equiv \begin{cases} G_t & t \leq \tau \\ G_t\left(I - P^{*\tau}_t\right) & t \geq \tau + 1 \end{cases}$$

Again consider the partition for $t \geq \tau + 1$

$$a_t = a^{*\tau}_t + a^{*\tau\perp}_t$$

$$x_t = x^{\tau*}_t + x^{*\tau\perp}_t$$

By construction of $H^{*\tau}_{\tau+1}$, it was the case that $x^{*\tau}_{\tau+1} = 0$, but that need not have been the case for $x^*_t = 0$ for $t > \tau + 1$. The variation was not coming from w_{t-1} or η_{t-1} as those are both 0. Instead it, was coming from potentially non-zero $a^{*\tau}_{t-1}$.

The newly defined $\tilde{a}_t$ has the property, however, that $\tilde{a}^{*\tau}_t = 0$ for all $t \geq \tau + 1$. Because $\tilde{x}^{*\tau}_{\tau+1} = x^{*\tau}_{\tau+1} = 0$, and $\tilde{a}^{*\tau}_s = w^{*\tau}_s = 0$ for all $\tau + 1 \leq s \leq t$, $\tilde{x}^{*\tau}_t = 0$. Again, recall that $h^{*\tau}_{t|\tau} = h^{*\tau}_t$, and $w_{t|\tau} = \eta_{t|\tau} = 0$. So, under the new FHS $\tilde{a}^{*\tau}_t = 0$ and $\tilde{x}^{*\tau}_t = 0$. This meets the Forward-Looking Constraints.

The FHS is designed so that it does not change outcomes in the orthogonal complement. $\tilde{a}^{*\perp\tau}_t = \left[\tilde{G}_t \quad I_{N_a} \quad 0\right] h^{*\perp\tau}_{t+1}$. Note that $\eta^{*\tau\perp}_{t+1} = \eta_{t+1}$. Also, $\tilde{G}_t h^{*\tau\perp}_t = G_t\left(I - P^{\tau*}_t\right) h^{*\tau\perp}_t = G_t h^{*\tau\perp}_t$. Together these show that $\tilde{a}^{*\tau\perp}_t = a^{*\tau\perp}_t$.

As $H_t^{*\tau}$ has no prediction errors, it does not affect $z_t - z_{t|t-1}$. Therefore, information updating will be exactly as before. Because actions are the same, and information updating is the same, the Forward-Looking Constraints will hold on $H_t^{*\tau\perp}$. Therefore, the Forward-Looking Constraints are met for all $t \geq \tau + 1$. And we have

$$\tilde{a}_t = a_t^{*\tau\perp}$$

$$\tilde{x}_t = x_t^{*\tau\perp}$$

so the variance will be weakly lower, as will losses

$$\text{Var}\left(\begin{bmatrix} a_t \\ x_t \end{bmatrix}\right) \geq \text{Var}\left(\begin{bmatrix} \tilde{a}_t \\ \tilde{x}_t \end{bmatrix}\right)$$

$$\left\langle L, \text{Var}\left(\begin{bmatrix} a_t \\ x_t \end{bmatrix}\right)\right\rangle_F \geq \left\langle L, \text{Var}\left(\begin{bmatrix} \tilde{a}_t \\ \tilde{x}_t \end{bmatrix}\right)\right\rangle_F$$

□

To summarize, if $a_{t+1|t}$ has variance beyond $x_{t|t}$ and $a_{t|t}$, then we can weakly reduce losses while maintaining the Forward-Looking Constraints by mapping that additional variance to 0. Put another way, $a_{t+1|t}$ has two potential uses: (i) to mitigate the period t Forward-Looking Constraint in response to $x_{t|t}$, $a_{t|t}$, or $x_{t+1|t}$; or (ii) to respond to $x_{t+1|t}$ for future losses or Forward-Looking Constraints. However, all of these depend on $x_{t|t}$, $a_{t|t}$, or $x_{t+1|t}$, and $x_{t+1|t} = A \begin{bmatrix} x_{t|t} \\ a_{t|t} \end{bmatrix}$, so the span of $\left(x_{t|t}, a_{t|t}\right)$ suffices. Additionally, any variance of $a_{t+1|t}$ cannot be a surprise, so it cannot change the information effect of future periods. In conclusion, the variance of $a_{t+1|t}$ beyond the mentioned span cannot help, and only serves to potentially increase losses.

Corollary 14. *There exists an optimal FHS* $\{G_t, \Sigma_t^\eta\}_{t=0}^\infty \in \mathcal{FH}$ *with the Span Property.*

Proof. Start with an optimal FHS, iterate forward through time weakly improving it at each time t where it does not have Span Property using Proposition 13. Once finished, it will have the Span Property for all t. Because the original FHS was an optimal one, losses will not decrease, but the proposition shows that they will not increase either. The fully improved FHS will also be an optimal one. □

This corollary is crucial to the proof because it allows us to restrict our analysis to optimal FHS with the Span Property. We will see that this subset of FHS have losses that can be precisely matched by a CES. Appendix A.4.2 shows an example of an FHS without the Span Property, how a CES could not have matched its covariances, and how Proposition 13 would weakly improve it to one with the Span Property.

Lemma 15. *Let a_t be defined by either an FHS or CES with the Span Property at time $t \geq 0$, i.e. $a_{t+1|t} = G_t^c \begin{bmatrix} x_{t|t} \\ a_{t|t} \end{bmatrix}$ for some $N_a \times N_{xa}$ G_t^c. Further define the $N_y \equiv 2N_x + N_a$ element augmented state y_t be defined as below with distribution Σ_t^y*

$$y_t \equiv \begin{bmatrix} x_t \\ x_{t|t-1} \\ a_{t|t-1} \end{bmatrix} \sim N\left(0, \Sigma_t^y\right)$$

Then, three things can be calculated based on the definitions above and $Var\left(\begin{bmatrix} y_t \\ a_t \end{bmatrix}\right)$: (i) the expected losses at period t, $E\left\{L_t\right\}$; (ii) whether the Forward-Looking Constraint is met at time t; and (iii) the next augmented state, y_{t+1}, and its, Σ_{t+1}^y.

Proof. First, define three convenience $N_{xa} \times 2N_{xa}$ matrices[9] that select the true values, the private sector's prior estimate, and the private sector's prediction error for $\begin{bmatrix} x_t \\ a_t \end{bmatrix}$

[9] These matrices have no time t subscript because they are constant.

from $\begin{bmatrix} y_t \\ a_t \end{bmatrix}$,

$$e^{xa} \equiv \begin{bmatrix} I & 0 & 0 & 0 \\ 0 & 0 & 0 & I \end{bmatrix}$$

$$\begin{bmatrix} x_t \\ a_t \end{bmatrix} = e^{xa} \begin{bmatrix} y_t \\ a_t \end{bmatrix}$$

$$e^{xa}_{|-1} \equiv \begin{bmatrix} 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \end{bmatrix}$$

$$\begin{bmatrix} x_{t|t-1} \\ a_{t|t-1} \end{bmatrix} = e^{xa}_{|-1} \begin{bmatrix} y_t \\ a_t \end{bmatrix}$$

$$e^{xa|ps}_{|-1} \equiv e^{xa} - e^{xa}_{|-1}$$

$$\begin{bmatrix} x_t - x_{t|t-1} \\ a_t - a_{t|t-1} \end{bmatrix} = e^{xa|ps}_{|-1} \begin{bmatrix} y_t \\ a_t \end{bmatrix}$$

(i) Losses: Observe,

$$\text{Var}\left(\begin{bmatrix} x_t \\ a_t \end{bmatrix}\right) = e^{xa}\text{Var}\left(\begin{bmatrix} y_t \\ a_t \end{bmatrix}\right)(e^{xa})^T$$

and then by Lemma 2.4.1,

$$\begin{aligned} L_t &= \left\langle L, \text{Var}\left(\begin{bmatrix} x_t \\ a_t \end{bmatrix}\right)\right\rangle \\ &= \left\langle L, e^{xa}\text{Var}\left(\begin{bmatrix} y_t \\ a_t \end{bmatrix}\right)(e^{xa})^T\right\rangle \end{aligned}$$

(ii) Constraint: The Forward-Looking Constraint at time t is

$$0 = D\begin{bmatrix} x_t \\ a_t \end{bmatrix} + J\begin{bmatrix} x_{t+1|t} \\ a_{t+1|t} \end{bmatrix}$$

Observe that from (2.6) and the assumption we have

$$\begin{bmatrix} x_{t+1|t} \\ a_{t+1|t} \end{bmatrix} = \begin{bmatrix} A \\ G^c_t \end{bmatrix} \begin{bmatrix} x_{t|t} \\ a_{t|t} \end{bmatrix}$$

Here are the steps to calculate $x_{t|t}$ and $a_{t|t}$. First, calculate the variance of the private-sector prediction errors,

$$\begin{bmatrix} x_t - x_{t|t-1} \\ a_t - a_{t|t-1} \end{bmatrix} = e^{xa|ps}_{|-1} \begin{bmatrix} y_t \\ a_t \end{bmatrix} \sim N\left(0, \Sigma^{xa|ps}_{t|t-1}\right)$$

$$\Sigma^{xa|ps}_{t|t-1} = e^{xa|ps}_{|-1} \text{Var}\left(\begin{bmatrix} y_t \\ a_t \end{bmatrix}\right) \left(e^{xa|ps}_{|-1}\right)^T$$

$z_t = C \begin{bmatrix} x_t \\ a_t \end{bmatrix}$, so the prediction error for z_t and its variance are

$$z_t - z_{t|t-1} = C \begin{bmatrix} x_t - x_{t|t-1} \\ a_t - a_{t|t-1} \end{bmatrix} \sim N\left(0, \Sigma^{z|ps}_{t|t-1}\right)$$

$$\Sigma^{z|ps}_{t|t-1} = C \Sigma^{xa|ps}_{t|t-1} C^T$$

Appendix A.2 shows that an optimal Kalman update is

$$K^{xa}_t \equiv \Sigma^{xa|ps}_{t|t-1} C^T \left(\Sigma^{z|ps}_{t|t-1}\right)^+$$

and

$$\begin{bmatrix} x_{t|t} \\ a_{t|t} \end{bmatrix} = \begin{bmatrix} x_{t|t-1} \\ a_{t|t-1} \end{bmatrix} + K^{xa}_t \left(z_t - z_{t|t-1}\right)$$

$$= \left(e^{xa}_{|-1} + K^{xa}_t C e^{xa|ps}_{|-1}\right) \begin{bmatrix} y_t \\ a_t \end{bmatrix}$$

So the Forward-Looking Constraint can be checked by whether the following equality is met for all realizable y_t, a_t, in the same fashion as Lemma 24 in Appendix A.1,

$$0 = \left(De^{xa} + J \begin{bmatrix} A \\ G^c_t \end{bmatrix} \left(e^{xa}_{|-1} + K^{xa}_t C e^{xa|ps}_{|-1}\right)\right) \begin{bmatrix} y_t \\ a_t \end{bmatrix}$$

(iii) The next augmented state, y_{t+1}, and its, Σ^y_{t+1}

This flows naturally from things we have already calculated. Define

$$A^{xa}_{t+1|t} \equiv \begin{bmatrix} A \\ G^c_t \end{bmatrix} \left(e^{xa}_{|-1} + K^{xa}_t C e^{xa|ps}_{|-1}\right)$$

$$\begin{bmatrix} x_{t+1|t} \\ a_{t+1|t} \end{bmatrix} = A^{xa}_{t+1|t} \begin{bmatrix} y_t \\ a_t \end{bmatrix}$$

Now observe

$$y_{t+1} \equiv \begin{bmatrix} x_{t+1} \\ x_{t+1|t} \\ a_{t+1|t} \end{bmatrix} = \begin{bmatrix} Ae^{xa} \\ A^{xa}_{t+1|t} \end{bmatrix} \begin{bmatrix} y_t \\ a_t \end{bmatrix} + \begin{bmatrix} B \\ 0 \\ 0 \end{bmatrix} w_{t+1}$$

$$\sim N\left(0, \Sigma^y_{t+1}\right)$$

$$\Sigma^y_{t+1} = \begin{bmatrix} Ae^{xa} \\ A^{xa}_{t+1|t} \end{bmatrix} \mathrm{Var}\left(\begin{bmatrix} y_t \\ a_t \end{bmatrix}\right) \left(\begin{bmatrix} Ae^{xa} \\ A^{xa}_{t+1|t} \end{bmatrix}\right)^T + \begin{bmatrix} B \\ 0 \\ 0 \end{bmatrix} \left(\begin{bmatrix} B \\ 0 \\ 0 \end{bmatrix}\right)^T$$

□

Lemma 15 shows that for an FHS or CES with the Span Property at t, Σ^y_t is a sufficient statistic for calculating the concurrent effects of the distribution of a_t. From the joint distribution of (y_t, a_t), it is possible to calculate the expected losses and whether the Forward-Looking Constraint is met.

Proposition 16. *Let FHS $\{G_t, \Sigma^\eta_t\}_{t=0}^\infty \in \mathcal{FH}$ with the Span Property. There exists a CES $\left\{G^e_t, G^c_t, \Sigma^{\tilde{\eta}}_t\right\}_{t=0}^\infty \in \mathcal{CE}$ with the same expected losses at every period.*

Proof. This proof is done by constructing the CES recursively. Because FHS has the Span Property, there exists a sequence of $N_a \times N_{xa}$ matrices $\{G^c_t\}_{t=0}^\infty$ such that $a_{t+1|t} = G^c_t \begin{bmatrix} x_{t|t} \\ a_{t|t} \end{bmatrix}$ for all t. These will be the ones used in the CES.

Based on Lemma 15, it suffices to show that Σ^y_0 is the same, and that for all $t \geq 0$ FHS and CES have the same $\mathrm{Var}\left(\begin{bmatrix} y_t \\ a_t \end{bmatrix}\right)$.

Under any FHS or CES,

$$\Sigma^y_0 = \begin{bmatrix} \Sigma^x_0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

as the private sector has no information before $t = 0$.

The inductive step is to show that given Σ_t^y is the same under both sequences, we can choose G_t^e and $\Sigma_t^{\tilde{\eta}}$ of the CES so that $\text{Var}\left(\begin{bmatrix} y_t \\ a_t \end{bmatrix}\right)$ is the same as under the FHS. It is potentially the case that the $\tilde{\eta}_t$ in the CES will have a different variance than the η_t in the FHSP. See Appendix A.4.1 for such an example.

From the FHS calculate the distribution of the random vector of private-sector prediction errors, with $e_{|-1}^{xa|ps}$ as defined in the lemma,

$$\begin{bmatrix} x_t - x_{t|t-1} \\ a_t - a_{t|t-1} \end{bmatrix} = e_{|-1}^{xa|ps} \begin{bmatrix} y_t \\ a_t \end{bmatrix} \sim N\left(0, \begin{bmatrix} \Sigma_{t|t-1}^{x|ps} & \Sigma_{t|t-1}^{xa|ps} \\ \Sigma_{t|t-1}^{ax|ps} & \Sigma_{t|t-1}^{a|ps} \end{bmatrix}\right)$$

$$\begin{bmatrix} \Sigma_{t|t-1}^{x|ps} & \Sigma_{t|t-1}^{xa|ps} \\ \Sigma_{t|t-1}^{ax|ps} & \Sigma_{t|t-1}^{a|ps} \end{bmatrix} = e_{|-1}^{xa|ps} \text{Var}\left(\begin{bmatrix} y_t \\ a_t \end{bmatrix}\right) \left(e_{|-1}^{xa|ps}\right)^T$$

Appendix A.3 shows that choosing G_t^e and $\Sigma_t^{\tilde{\eta}}$ as below will replicate the distribution of private-sector prediction errors,[10]

$$G_t^e \equiv \Sigma_{t|t-1}^{ax|ps} \left(\Sigma_{t|t-1}^{x|ps}\right)^+$$

$$\Sigma_t^{\tilde{\eta}} \equiv \Sigma_{t|t-1}^{a|ps} - \Sigma_{t|t-1}^{ax|ps} \left(\Sigma_{t|t-1}^{x|ps}\right)^+ \Sigma_{t|t-1}^{xa|ps}$$

$$\left(a_t - a_{t|t-1}\right) = G_t^e \left(x_t - x_{t|t-1}\right) + \tilde{\eta}_t$$

$$\tilde{\eta}_t \sim N\left(0, \Sigma_t^{\tilde{\eta}}\right)$$

By replicating the prediction error covariance, we see that the definition of a_t in CES, $a_t = a_{t|t-1} + G_t^e \left(x_t - x_{t|t-1}\right) + \tilde{\eta}_t$ replicates $\text{Var}\left(\begin{bmatrix} y_t \\ a_t \end{bmatrix}\right)$ from the FHS.

As Σ_t^y and $\text{Var}\left(\begin{bmatrix} y_t \\ a_t \end{bmatrix}\right)$ are the same, Lemma 15 shows that the losses will be the same, the CES will meet the Forward-Looking Constraint, and Σ_{t+1}^y will be the

[10] This is where allowing η_t in the definition for a_t is extremely helpful. If CES did not have η_t, we'd have to limit our analysis to FHS for which $\Sigma_{t|t-1}^{a|ps} = \Sigma_{t|t-1}^{ax|ps} \left(\Sigma_{t|t-1}^{x|ps}\right)^+ \Sigma_{t|t-1}^{xa|ps}$, i.e. there is no variation in the private-sector prediction error for a_t beyond its covariance with the private-sector prediction error for x_t.

same. □

The above theorem shows that the losses of an FHS from $\mathcal{FH}$ with the Span Property can be matched exactly with a CES from $\mathcal{CE}$. It uses the $\{G_t^c\}_{t=0}^{\infty}$ from the FHS having the Span Property, and recursively chooses $\left\{G_t^e, \Sigma_t^{\tilde{\eta}}\right\}_{t=0}^{\infty}$ to match the joint covariance of y_t and a_t at every period t. Crucially, this implies that the losses for an optimal FHS with the Span Property can be attained by a CES as well.

The next proposition shows that we can go in the opposite direction as well. Every CES has a corresponding FHS.

Proposition 17. *Given a CES* $\{G_t^e, G_t^c, \Sigma_t^{\eta}\}_{t=0}^{\infty} \in \mathcal{CS}$*, there exists an FHS* $\{G_t, \Sigma_t^{\eta}\}_{t=0}^{\infty} \in \mathcal{FH}$ *that has the same* a_t *at all times.*

Proof. Let CES $\{G_t^e, G_t^c, \Sigma_t^{\eta}\}_{t=0}^{\infty} \in \mathcal{CS}$, meaning that for $t \geq 0$

$$a_0 = G_0^e x_0 + \eta_0$$

$$a_{t+1} = G_t^c \begin{bmatrix} x_{t|t} \\ a_{t|t} \end{bmatrix} + G_{t+1}^e \left(x_t - x_{t+1|t}\right) + \eta_{t+1}$$

$$\eta_t \sim N\left(0, \Sigma_t^{\eta}\right)$$

We wish to calculate $\{G_t\}_{t=0}^{\infty}$ so that $a_t = G_t h_t + \eta_t$. (We will be keeping Σ_t^{η} the same.)

We do this inductively, accumulating two utility matrices, $P_{t-1|t-1}^{xa}$ and φ_t, such that

$$\begin{bmatrix} x_{t-1|t-1} \\ a_{t-1|t-1} \end{bmatrix} = P_{t-1|t-1}^{xa} h_t$$

$$x_t = \varphi_t h_t$$

As a reminder

$$h_0 = x_0 \sim N\left(0, \Sigma_0^x\right)$$

$$h_{t+1} = \begin{bmatrix} h_t \\ \eta_t \\ w_{t+1} \end{bmatrix} \sim N\left(0, \Sigma_{t+1}^h\right)$$

$$\Sigma_{t+1}^h = \begin{bmatrix} \Sigma_t^h & 0 & 0 \\ 0 & \Sigma_t^\eta & 0 \\ 0 & 0 & \Sigma^w \end{bmatrix}$$

We initialize the two utility matrices

$$P_{-1|-1}^{xa} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\varphi_0 = I_{N_x}$$

For $t = 0$

$$G_0 = G_0^e$$

$$a_0 = G_0 h_0 + \eta_0$$

Now for every $t \geq 1$, calculate

$$a_t = G_{t-1}^c \begin{bmatrix} x_{t-1|t-1} \\ a_{t-1|t-1} \end{bmatrix} + G_t^e \left(x_t - x_{t|t-1}\right) + \eta_t$$

$$x_{t|t-1} = A \begin{bmatrix} x_{t-1|t-1} \\ a_{t-1|t-1} \end{bmatrix}$$

$$G_t = G_{t-1}^c P_{t-1|t-1}^{xa} + G_t^e \left(\varphi_t - A P_{t-1|t-1}^{xa}\right)$$

$$a_t = G_t h_t + \eta_t$$

This is a crucial calculation for this step. Combining $P_{t-1|t-1}^{xa}$, φ_t, and A with choices G_{t-1}^c and G_t^e, we calculate the exact G_t that will have the same a_t as the CES. Σ_t^η remains the same in both the CES and FHS.

Now however, we need to do the more complicated calculation of $P_{t|t}^{xa}$ to be used next period. To do that we will have to calculate the Kalman updating, which will depend on the uncertainty in (x_t, a_t). Observe

$$\begin{bmatrix} x_{t|t-1} \\ a_{t|t-1} \end{bmatrix} = \begin{bmatrix} A \\ G^c_{t-1} \end{bmatrix} P^{xa}_{t-1|t-1} h_t$$

Define the utility matrix $P^{xa|ps}_{t|t-1}$, and error in projection onto h_t as,

$$P^{xa|ps}_{t|t-1} \equiv \begin{bmatrix} I \\ G^e_t \end{bmatrix} \left(\varphi_t - A P^{xa}_{t-1|t-1}\right)$$

$$\begin{bmatrix} x_t - x_{t|t-1} \\ a_t - a_{t|t-1} \end{bmatrix} = P^{xa|ps}_{t|t-1} h_t + \begin{bmatrix} 0 \\ \eta_t \end{bmatrix}$$

The covariance of the error is then

$$\Sigma^{xa|ps}_{t|t-1} = P^{xa|ps}_{t|t-1} \Sigma^h_t \left(P^{xa|ps}_{t|t-1}\right)^T + \begin{bmatrix} 0 & 0 \\ 0 & \Sigma^\eta_t \end{bmatrix}$$

Then the Kalman update is the familiar

$$K^{xa}_t \equiv \Sigma^{xa|ps}_{t|t-1} C^T \left(C \Sigma^{xa|ps}_{t|t-1} C^T\right)^+$$

Now we can to bookkeeping to track the update

$$\begin{bmatrix} x_{t|t} \\ a_{t|t} \end{bmatrix} = \begin{bmatrix} x_{t|t-1} \\ a_{t|t-1} \end{bmatrix} + K^{xa}_t \left(z_t - z_{t|t-1}\right)$$

$$= (I - K^{xa}_t C) \begin{bmatrix} x_{t|t-1} \\ a_{t|t-1} \end{bmatrix} + K^{xa}_t C \begin{bmatrix} x_t \\ a_t \end{bmatrix}$$

$$= (I - K^{xa}_t C) \begin{bmatrix} A \\ G^c_t \end{bmatrix} P^{xa}_{t-1|t-1} h_t + K^{xa}_t C \begin{bmatrix} \varphi_t \\ G_t \end{bmatrix} h_t + K^{xa}_t C \begin{bmatrix} 0 \\ I_{N_a} \end{bmatrix} \eta_t$$

Define the first part that gets applied to h_t, and then construct $P^{xa}_{t|t}$

$$P^{xa|h}_{t|t} \equiv (I - K^{xa}_t C) \begin{bmatrix} A \\ G^c_t \end{bmatrix} P^{xa}_{t-1|t-1} + K^{xa}_t C \begin{bmatrix} \varphi_t \\ G_t \end{bmatrix}$$

$$P^{xa}_{t|t} = \begin{bmatrix} P^{xa|h}_{t|t} & K^{xa}_t C \begin{bmatrix} 0 \\ I_{N_a} \end{bmatrix} & 0_{N_{xa} \times N_w} \end{bmatrix}$$

$$\begin{bmatrix} x_{t|t} \\ a_{t|t} \end{bmatrix} = P^{xa}_{t|t} \begin{bmatrix} h_t \\ \eta_t \\ w_{t+1} \end{bmatrix} = P^{xa}_{t|t} h_{t+1}$$

And then to construct φ_{t+1} in the same fashion as in Lemma 9,

$$\varphi_{t+1} = \begin{bmatrix} A \begin{bmatrix} \varphi_t & 0 \\ G_t & I_{N_a} \end{bmatrix} & B \end{bmatrix}$$

$$x_{t+1} = \varphi_{t+1} h_{t+1}$$

□

Theorem 18. *The value functions for the FHSP and CESP are equal,* $V(\Sigma_0^x) = W(\Sigma_0^x)$*, and an optimal CES* $\{G_t^e, G_t^c, \Sigma_t^\eta\}_{t=0}^\infty \in \mathcal{CE}$ *has a corresponding optimal FHS* $\{G_t, \Sigma_t^\eta\}_{t=0}^\infty \in \mathcal{FH}$.

Proof. Corollary 14 says there exists an optimal FHS with the Span Property. Proposition 16 shows that the losses from such a FHS can be matched with a CES. Therefore, the optimal FHS cannot outperform the optimal CES, $V(\Sigma_0^x) \geq W(\Sigma_0^x)$. Proposition 17 says that any CES can be translated into the equivalent FHS. Therefore, the optimal CES cannot outperform an optimal FHS, $V(\Sigma_0^x) \leq W(\Sigma_0^x)$, and together $V(\Sigma_0^x) = W(\Sigma_0^x)$. Additionally, if CES $\{G_t^e, G_t^c, \Sigma_t^\eta\}_{t=0}^\infty \in \mathcal{CE}$ is optimal, then its corresponding FHS is also optimal. □

Theorem 19. *The value function of the CESP and the value function of the VBP are equal when* U *is initialized with* Σ_0^y*,*

$$W(\Sigma_0^x) = U(\Sigma_0^y)$$

$$\Sigma_0^y = \begin{bmatrix} \Sigma_0^x & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The optimal VBP functions yield an optimal CES,

$$\{G_t^e(\Sigma_t^y), G_t^c(\Sigma_t^y), \Sigma_t^\eta(\Sigma_t^y)\}_{t=0}^\infty$$

for

$$\Sigma_0^y = \begin{bmatrix} \Sigma_0^x & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\Sigma_{t+1}^y = \Sigma_n^y \left(\Sigma_t^y\right)$$

Proof. Lemma 15 shows that we can calculate $W\left(\Sigma_0^x\right)$ by initializing Σ_0^y and plugging in the optimal choices for the CESP as we go to calculate every L_t and then perform the discounted sum. Every one of those choices is a valid option for the VBP, and the discounted sum is the same as the one performed inside the VBP. Therefore, $W\left(\Sigma_0^x\right) \geq U\left(\Sigma_0^y\right)$.

Note that by construction, the Forward-Looking Constraint inside the definition of $\mathcal{CE}$ is

$$0 = D \begin{bmatrix} x_t \\ a_t \end{bmatrix} + J \begin{bmatrix} A \\ G_t^c \end{bmatrix} \begin{bmatrix} x_{t|t} \\ a_{t|t} \end{bmatrix}$$

Also, by constructing a_{t+1} in the CES fashion,

$$E\left\{ \begin{bmatrix} x_t \\ a_t \\ x_{t+1} \\ a_{t+1} \end{bmatrix} | I_t \right\} = E\left\{ \begin{bmatrix} x_t \\ a_t \\ x_{t+1} \\ a_{t+1} \end{bmatrix} | z_t, x_{t-1|t-1}, a_{t-1|t-1} \right\}$$

as $\left(z_t, x_{t-1|t-1}, a_{t-1|t-1}\right)$ contain as much information about the estimated variables as $\{z_\tau | \tau \leq t\}$.

Let $G^e\left(\Sigma^y\right)$, $G^c\left(\Sigma^y\right)$, $\Sigma^\eta\left(\Sigma^y\right)$ be the optimal choice functions for the VBP. We can generate a CES from $\mathcal{CE}$ by calculating for all $t \geq 0$

$$G_t^e = G^e\left(\Sigma_t^y\right)$$

$$G_t^c = G^c\left(\Sigma_t^y\right)$$

$$\Sigma_t^\eta = \Sigma^\eta\left(\Sigma_t^y\right)$$

$$\Sigma_{t+1}^y = \Sigma_n^y\left(\Sigma_t^y\right)$$

Thus, $W\left(\Sigma_0^x\right) \leq U\left(\Sigma_0^y\right)$. Together, $W\left(\Sigma_0^x\right) = U\left(\Sigma_0^y\right)$. □

The equivalence of CESP and VBP is because each period they have the same choice sets and constraints. When the policy is constructed in the y_t, G^c, G^e fashion, Σ^y forms a sufficient statistic for calculating losses and tracking the private sector's information. The difference between the two formulations is how the optimization takes place. The CESP optimizes all the choices concurrently to minimize the expectation of a discounted sum of losses. The VBP uses a value function to minimize the choices each period, and then, in effect, perform a discounted sum of expectations. But both problems are solved by the policymaker before any shocks have realized.

2.5 New Keynesian Example

As an application I repeat the analysis Mertens (2016), but solving for commitment behavior instead of discretionary behavior.

This is a textbook New Keynesian monetary model with two small modifications. The log–linearized Phillips curve is

$$\pi_t = \beta\pi_{t+1|t} + \kappa g_t$$

where π_t is inflation, $\beta \in [0, 1)$ is the common discount factor, κ is a reduced-form slope that depends on the microfoundations, and g_t is the output gap.[11] As above, $\pi_{t+1|t} = E\{\pi_{t+1}|I_t^{ps}\}$, where I_t^{ps} is the private sector's information set. I follow his numerical exercise, in setting equal weights for the central bank on the inflation and the output gap target deviation

$$L_t = \pi_t + (g_t - \overline{g}_t)^2$$

The unconditional expected losses at time 0 are

$$L = E\left[\sum_{t=0}^{\infty} \beta^t L_t\right]$$

[11] I use unconventional variable g for the output gap, because x and y are used in the general model.

The output gap target, $\overline{g}_t$ is time varying. I follow Mertens (2016) and Cukierman and Meltzer (1986), interpreting the time variation in this target as changing central-bank preferences. In the financial press, there is discussion of different Federal Reserve leaders as either hawkish or dovish, and I interpret this as suggesting that there is uncertainty about the Federal Reserve's target for the output gap. Tang (2015) provides an alternative interpretation, as "exogenous variation in the wedge between the efficient and flexible-price levels of output."

Again following Mertens (2016), the output target has an autoregressive component and an uncorrelated component

$$\overline{g}_t = \gamma_t + \varepsilon_t$$

$$\gamma_t = \rho\gamma_{t-1} + \nu_t$$

$$\begin{bmatrix} \nu_t \\ \varepsilon_t \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma_\nu^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{bmatrix}\right)$$

for constant $\rho \in (0, 1)$, and variances σ_ε^2, σ_ν^2.

Following Mertens, we consider three informational setups, Full Information, where the private sector observes γ_t and ε_t; Lagged Information, where the private sector observes γ_{t-1} and π_t, (which is enough to infer ε_{t-1}); and finally Dynamic Information, where the private sector observes only π_t.[12] In the graphs I also include the Discretionary version of Dynamic information for comparison.

Like in the two period model, we can ensure that the Forward-Looking Constraint is met by using using the matrix choices to determine π_t and $\pi_{t+1|t}$, and then using

[12] In the two-period model of Section 2.2, Lagged Information and Dynamic Information would be identical.

those to determine g_t,

$$\pi_t = \pi_{t|t-1} + G_t^e \left(\begin{bmatrix} \gamma_t \\ \varepsilon_t \end{bmatrix} - \begin{bmatrix} \gamma_{t|t-1} \\ 0 \end{bmatrix} \right)$$

$$\pi_{t+1|t} = G_t^c \begin{bmatrix} \gamma_{t|t} \\ \varepsilon_{t|t} \\ \pi_t \end{bmatrix}$$

$$g_t = \pi_t - \pi_{t+1|t}$$

with π_t instead of $\pi_{t|t}$ in the second equation because π_t is always in the private-sector information set.

The setup is slightly different for Lagged Information, as there are three elements in x_t now,

$$\begin{bmatrix} \gamma_{t+1} \\ \varepsilon_{t+1} \\ \gamma_t \end{bmatrix} = \begin{bmatrix} \rho & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \gamma_t \\ \varepsilon_t \\ \gamma_{t-1} \end{bmatrix} + \begin{bmatrix} \sigma_\nu & 0 \\ 0 & \sigma_\varepsilon \\ 0 & 0 \end{bmatrix} w_{t+1}$$

and

$$\pi_t = \pi_{t|t-1} + G_t^{e,LI} \left(\begin{bmatrix} \gamma_t \\ \varepsilon_t \\ \gamma_{t-1} \end{bmatrix} - \begin{bmatrix} \gamma_{t|t-1} \\ 0 \\ \gamma_{t-1|t-1} \end{bmatrix} \right)$$

$$\pi_{t+1|t} = G_t^{c,LI} \begin{bmatrix} \gamma_{t|t} \\ \varepsilon_{t|t} \\ \gamma_{t-1} \\ \pi_t \end{bmatrix}$$

$$g_t = \pi_t - \pi_{t+1|t}$$

with γ_{t-1} and π_t instead of $\gamma_{t-1|t}$ and $\pi_{t|t}$ because $\{\gamma_{t-1}, \pi_t\} \in I_t^{ps,LI}$.

We solve for the steady state by using the VBR formulation

$$U(\Sigma^y) = \min_{G^e, G^c} E\{L_c\} + \beta U(\Sigma_n^y)$$

where for optimal choices G^e and G^c yield $\Sigma_n^y = \Sigma^y$.

2.5.1 Finding the Steady State

I ensure that all possible $\left(G^e\left(\Sigma^y\right), G^c\left(\Sigma^y\right)\right)$ are in $\mathcal{FH}$ by using them to define π_t, and then using π_t to determine the g_t that meets the Forward-Looking Constraint,

$$\pi_c = \pi_{c|p} + G^e\left(\Sigma^y\right)\left(\begin{bmatrix}\gamma_c \\ \varepsilon_c\end{bmatrix} - \begin{bmatrix}\gamma_{c|p} \\ 0\end{bmatrix}\right)$$

$$\pi_{n|c} = G^c\left(\Sigma^y\right)\begin{bmatrix}\gamma_{c|c} \\ \varepsilon_{c|c} \\ \pi_c\end{bmatrix}$$

$$g_c = \pi_c - \pi_{n|c}$$

The Steady State Equilibrium is given by

$$U\left(\Sigma^{y*}\right) = \min_{G^e, G^c} E\left\{L_c\right\} + \beta U\left(\Sigma^y_n\right)$$

$$\Sigma^y_n\left(G^{e*}, G^{c*}\right) = \Sigma^{y*}$$

that is a distribution Σ^{y*} such that the optimal choices for G^e and G^c yield the same distribution for the next period. The key to finding the optimal G^e, G^c is calculating the $\partial U/\partial\Sigma^y$. By the envelope condition we can calculate that for a given G^e, G^c. The algorithm I use considers the half-vectorization that collects the lower triangular elements of Σ^y or Σ^y_n into a vector, denoted by function *vech*. It has $N_{\mathbf{y}} \equiv N_y\left(N_y+1\right)/2$ elements. From that vector, it is easy to reshape it back into a lower triangular matrix, and then use the transpose to create the full symmetric matrix.

Rephrasing in terms of *vech*,

$$\mathbf{y} \equiv vech\left(\Sigma^y\right)$$

$$\mathbf{y}^n \equiv vech\left(\Sigma^y_n\right)$$

$$U\left(\mathbf{y}\right) = \min_{G^e, G^c} E\left\{L_c\right\} + \beta U\left(\mathbf{y}^n\right)$$

The partial for a specific element of $\mathbf{y}_i$

$$\frac{\partial U}{\partial \mathbf{y}_i} = \frac{\partial E\left\{L_c\right\}\left(\Sigma^y, G^e, G^c\right)}{\partial \mathbf{y}_i} + \beta\left(\frac{\partial \mathbf{y}^n\left(\Sigma^y, G^e, G^c\right)}{\partial \mathbf{y}_i}\right)^T \frac{\partial U}{\partial \mathbf{y}}$$

where $\frac{\partial \mathbf{y}^n}{\partial \mathbf{y}_i}$ is the $N_{\mathbf{y}} \times 1$ vector of partials for the half-vectorization of the next period's distribution. In words, using the envelope condition, we know that the partial derivative of U with respect to a specific element of $\mathbf{y}$ (and therefore Σ^y) is equal to the sum of its direct effect on losses, and its direct effect on the distribution of the next period's distribution multiplied by the effects that has on the next period's expected losses. Collecting all the partials,

$$\frac{\partial U}{\partial \mathbf{y}} = \frac{\partial E\left\{L_c\right\}\left(\Sigma^y, G^e, G^c\right)}{\partial \mathbf{y}} + \beta J^T \frac{\partial U}{\partial \mathbf{y}}$$

where

$$J_{ij} = \frac{\partial \mathbf{y}_i^n\left(\Sigma^y, G^e, G^c\right)}{\partial \mathbf{y}_j}$$

represents the Jacobian of $\mathbf{y}^n$ with respect to $\mathbf{y}$. And we can rearrange

$$\frac{\partial U}{\partial \mathbf{y}} = \left(I - \beta J^T\right)^{-1} \frac{\partial E\left\{L_c\right\}\left(\Sigma^y, G^e, G^c\right)}{\partial \mathbf{y}}$$

Now we can state the first order condition for G^e and G^c,

$$0 = \frac{\partial E\left\{L_c\right\}}{\partial G^e} + \beta\left(\frac{\partial \mathbf{y}^n}{\partial G^e}\right)^T \frac{\partial U}{\partial \mathbf{y}}$$

$$0 = \frac{\partial E\left\{L_c\right\}}{\partial G^c} + \beta\left(\frac{\partial \mathbf{y}^n}{\partial G^c}\right)^T \frac{\partial U}{\partial \mathbf{y}}$$

So the procedure is as follows. Choose an initial $G^{e(0)}$ and $G^{c(0)}$.

The first step is to find the Σ^y implied by $\left(G^{e(i)}, G^{c(i)}\right)$. We can do this by iterating equation (A.4) from Appendix A.1, starting with $\Sigma_0^y = \begin{bmatrix} \Sigma_0^x & 0 \\ 0 & 0 \end{bmatrix}$

$$\Sigma_n^y = \begin{bmatrix} AG_c^{xa} & B \\ AP_{c|c}^{xa} & 0 \\ G^{c(i)}P_{c|c}^{xa} & 0 \end{bmatrix} \begin{bmatrix} \Sigma^y & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \left(\begin{bmatrix} AG_c^{xa} & B \\ AP_{c|c}^{xa} & 0 \\ G^{c(i)}P_{c|c}^{xa} & 0 \end{bmatrix} \right)^T$$

where G_c^{xa} is defined in (A.1), and is a function of $G^{e(i)}$, and $P_{c|c}^{xa}$ is defined in equation (A.3), and depends on Σ^y and $G^{e(i)}$.

Once we have found the steady state distribution, $\Sigma^{y(i)}$, extract it's corresponding $\mathbf{y}^{(i)} \equiv vech\left(\Sigma^{y(i)}\right)$ we can use we use it to calculate the partial, $\frac{\partial U}{\partial \mathbf{y}}$, and from that we can calculate the FOC for G^e and G^c. Then, the question is how to generate the next $\left(G^{e(i+1)}, G^{c(i+1)}\right)$. What I have found works to approximate U in the recursive call as

$$\tilde{U}^{(i)}\left(\tilde{\mathbf{y}}\right) = \left(\frac{\partial U}{\partial \mathbf{y}}\left(\mathbf{y}^{(i)}\right)\right)^T \left(\tilde{\mathbf{y}} - \mathbf{y}^{(i)}\right) + \left(\tilde{\mathbf{y}} - \mathbf{y}^{(i)}\right)^T \tilde{H}_U^{(i)} \left(\tilde{\mathbf{y}} - \mathbf{y}^{(i)}\right)$$

where $\tilde{H}_U$ is an approximation of the Hessian of U that holds G^e and G^c fixed at their previous values,

$$\tilde{H}_U^{(i)} = \frac{\partial}{\partial \mathbf{y}^T}\left(\frac{\partial U\left(\mathbf{y}^{(i)}, G^{e(i)}, G^{c(i)}\right)}{\partial \mathbf{y}}\right)$$

Because this is a second derivative, the envelope condition no longer applies, so this is not the true Hessian of U. At the steady state equilibrium, $\tilde{\mathbf{y}}$ will approach $\mathbf{y}^{(i)}$, so the inaccurate $\tilde{H}_U$ will not mater compared to the accurate partial, $\left(\frac{\partial U}{\partial \mathbf{y}}\right)^T$.

Using $\tilde{H}_U$, we calculate the next iteration of G^e and G^c by

$$G^{e(i+1)}, G^{c(i+1)} \equiv \arg\min_{G^e, G^c}\left(E\left\{L_c\right\} + \beta\tilde{U}\left(\mathbf{y}^n\left(\mathbf{y}, G^e, G^c\right)\right)\right)$$

To find the optimal values for Discretion, I implement Mertens (2016, Online

TABLE 2.1: Optimal Policies G^e and G^c across Information Structures

	FI	LI	DI	Discretion
G^e_γ	0.58	0.31	0.29	0.21
G^e_ε	0.38	0.54	0.67	0.44
$G^e_{\gamma-1}$		0.25		
G^c_γ	-0.06	-0.15	-0.06	0.24
G^c_ε	-0.38	-0.33	-0.38	
$G^c_{\gamma-1}$		0.08		
G^c_π	0.38	0.38	0.38	

Appendix) method. It yields an optimal

$$\pi_t^{disc} = G^{disc} \begin{bmatrix} \gamma_t \\ \varepsilon_t \\ \gamma_{t|t-1} \end{bmatrix}$$

$$= \pi_{t|t-1}^{disc} + G^{disc} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \left(\begin{bmatrix} \gamma_t \\ \varepsilon_t \end{bmatrix} - \begin{bmatrix} \gamma_{t|t-1} \\ 0 \end{bmatrix} \right)$$

$$\pi_{t|t-1}^{disc} = G^{disc} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \rho \gamma_{t-1|t-1}$$

So, we can treat them as choosing the same constants, but optimizing G^c_γ from the perspective of the next period's policymaker,

$$G^{e,Disc} = G^{disc} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$$G^{c,Disc}_\gamma = \rho G^{disc} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

2.5.2 Results

In order to ease interpretability of tradeoffs between inflation and the output gap, I use $\kappa = 1$ for the numerical exercise. The other parameters are the same as used in Mertens, $\beta = 0.99$, $\rho = 0.9$, and unit shocks to γ_t and ε_t each period.

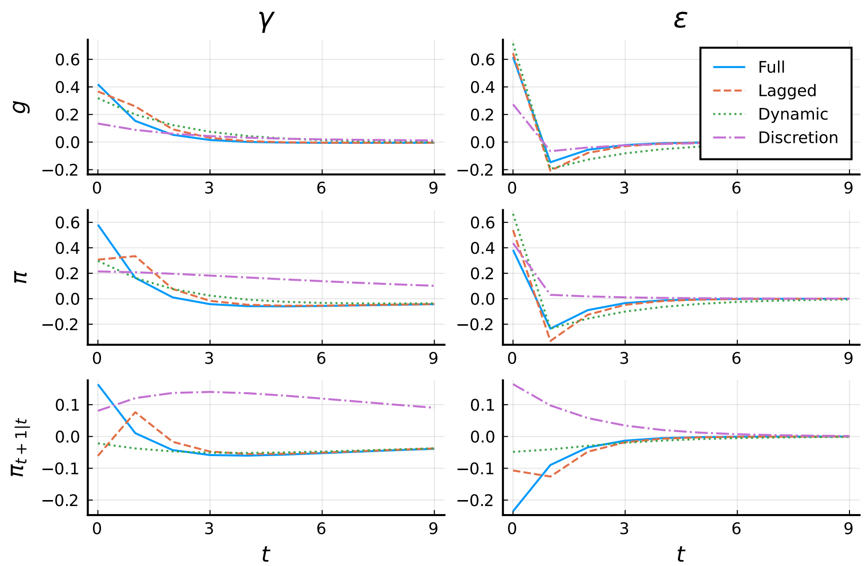

FIGURE 2.1: Impulse Response across Information Structures

The left column of Figure 2.1 shows the impulse responses to a unit shock to the persistent shock, γ_0, and the right shows to unit transitory shock, ε_0. The rows are the realized output gap, g_t, realized inflation, π_t, and the private sector's next period expected inflation, $\pi_{t+1|t}$.

Under Full Information, the $\pi_{t+1|t}$ graph is the π graph shifted to the left by one. It is easy to see how this represents the commitment made by the central bank to the private sector.

What is more subtle is how for the other two information sets, $\pi_{t+1|t}$ differs from π_{t+1}, but it is still a commitment followed through upon by the policymaker. The forward-looking commitment is with regard to the *private sector's information*, $\pi_{t+1|t} = G^c \begin{bmatrix} x_{t|t} \\ \pi_t \end{bmatrix}$. Because the private sector in the shock period has informational errors for both Lagged Information and Dynamic Information, the private sector's forecast will not equal the central bank's, $\pi_{t+1|t} \neq E\left\{\pi_{t+1}|I_t^{cb}\right\}$, where I_t^{cb} is the information set of the central bank.

For Lagged Information, the difference only exists in the first period. In the bottom left panel, the Lagged public sector expects $\pi_{1|0}^{LI,\gamma} = -0.06$, but the true value is actually $\pi_1^{LI,\gamma} = 0.33$. For the ε shock, the private sector overestimates inflation with $\pi_{1|0}^{LI,\varepsilon} = -0.11$ and $\pi_1^{LI,\varepsilon} = -0.33$. The central bank is still following through on its commitment, despite the inaccurate private-sector prediction because the difference comes from the private sector's prediction error with regard to γ_1

$$\pi_1 = \pi_{1|0} + G^{e,LI}\left(\begin{bmatrix}\gamma_1\\ \varepsilon_1\\ \gamma_0\end{bmatrix} - \begin{bmatrix}\gamma_{1|0}\\ 0\\ \gamma_{0|0}\end{bmatrix}\right)$$

Because the private sector has an inaccurate $\gamma_{0|0}$, $\pi_{1|0}$ differs from $E\left\{\pi_1|I_0^{cb}\right\}$. These values are shown in Figure 2.2. We can see that in period 1, for the Lagged Information private sector, the inflation prediction error is even higher than in the first period. It is higher the inflation prediction error in the first period, despite $\gamma_1 - \gamma_{1|0}$ being lower because inaccurate $\gamma_{0|0}$ affects the errors for both γ_1 and the lag element of x_1, γ_0. Thus in period 1 after $\gamma_0 = 1$ shock, the total weight on inaccurate $\gamma_{0|0}$ is $G_\gamma^e \rho \left(1 - \gamma_{0|0}\right) + G_{\gamma_{-1}}^e \left(1 - \gamma_{0|0}\right)$. After period 1, the Lagged Information private sector has accurate information and $\pi_{t+1|t}^{LI} = \pi_{t+1}^{LI}$ for $t \geq 1$.

Now we consider the impulse responses for Dynamic Information. The effects of a shock are much more persistent than the other two informational setups. For both shocks, output and inflation much more gradually return to trend. After the persistent shock Figure 2.2 shows how the private sector takes a very long time to update its beliefs. This is primarily driven by the fact that the central bank responds more aggressively to ε_t than $\left(\gamma_t - \gamma_{t|t-1}\right)$, 0.67 and 0.29 respectively (also visible as the difference in heights for π_0 on the graphs). In the two-period model, these numbers were 0.56 and 0.311.

The slow updating means after a γ shock, the Dynamic Information private sector is persistently underestimating inflation. Every period $\gamma_t > \gamma_{t|t-1}$, so $\pi_t > \pi_{t|t-1}$. And each period, the update from the surprise is almost equally attributed to $\gamma_{t|t} - \gamma_{t|t-1}$

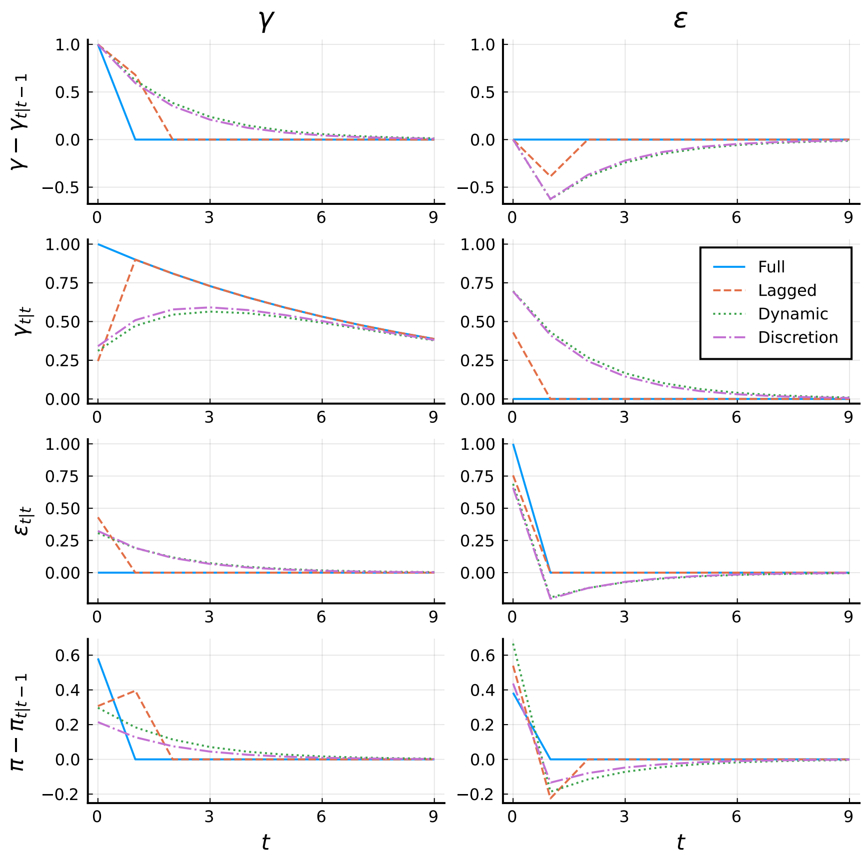

FIGURE 2.2: Private Sector Predictions and Errors

and $\varepsilon_{t|t}$, as seen most clearly in that $\gamma_{0|0}$ and $\varepsilon_{0|0}$ are almost equal for either shock, when $\gamma_{0|-1} = 0$. The approximate values are

$$\begin{bmatrix}\gamma_{t|t}\\ \varepsilon_{t|t}\end{bmatrix} = \begin{bmatrix}\gamma_{t|t-1}\\ 0\end{bmatrix} + \begin{bmatrix}1.037\\ 1.036\end{bmatrix}\left(\pi_t - \pi_{t|t-1}\right)$$

$$= \begin{bmatrix}\gamma_{t|t-1}\\ 0\end{bmatrix} + \begin{bmatrix}0.305 & 0.696\\ 0.304 & 0.695\end{bmatrix}\begin{bmatrix}\gamma_t - \gamma_{t|t-1}\\ \varepsilon_t\end{bmatrix}$$

The reason that the updating weights are close, despite the fact that the $G^{e,PI}$ weights differ is that the variance of the two quantities differ significantly

$$\mathrm{Var}\left(\gamma_t - \gamma_{t|t-1}\right) = 2.29$$

$$\mathrm{Var}\left(\varepsilon_t\right) = 1$$

and Kalman updating is a combination of uncertainty and observational weights

$$K^{DI} = \Sigma^{x|ps}_{t|t-1} \left(G^{e,DI}\right)^T \left(G^{e,DI}\Sigma^{x|ps}_{t|t-1} \left(G^{e,DI}\right)^T\right)^{-1}$$

$$= \begin{bmatrix} 1.037 \\ 1.036 \end{bmatrix}$$

Discretion has very similar information updating as Dynamic Information. This comes from the fact that the information in π_t determined by the ratio of G^e_γ to G^e_ε and is relatively close in both models

$$\frac{G^{e,DI}_\gamma}{G^{e,DI}_\varepsilon} = 0.45$$

$$\frac{G^{e,Disc}_\gamma}{G^{e,Disc}_\varepsilon} = 0.49$$

with the corresponding update weights being

$$\begin{bmatrix} \gamma_{t|t} \\ \varepsilon_{t|t} \end{bmatrix} = \begin{bmatrix} \gamma_{t|t-1} \\ 0 \end{bmatrix} + \begin{bmatrix} 1.59 \\ 1.51 \end{bmatrix} \left(\pi_t - \pi_{t|t-1}\right)$$

$$= \begin{bmatrix} \gamma_{t|t-1} \\ 0 \end{bmatrix} + \begin{bmatrix} 0.341 & 0.694 \\ 0.324 & 0.659 \end{bmatrix} \begin{bmatrix} \gamma_t - \gamma_{t|t-1} \\ \varepsilon_t \end{bmatrix}$$

The similarity of private-sector information is reflected in the first three rows of Figure 2.2.

However, the consequences for for outcomes are quite difference. The difference is driven primarily by expected inflation, shown in the third row of Figure 2.1. To some extent, after the γ shock at period 0, $\gamma_{1|0}$ is somewhat accurate. With regard $\gamma_{1|0}$, the DI policymaker with commitment can bind itself to

$$\pi^{DI,\gamma}_{1|0} = G^{c,DI} \begin{bmatrix} 0.305 \\ 0.304 \\ 0.297 \end{bmatrix} = -0.02$$

The Discretionary central bank must take the next period's $G^{c,Disc}$ as given

$$\pi^{Disc,\gamma}_{1|0} = G^{c,Disc}_\gamma \left[0.341\right] = 0.08$$

Expected inflation rises in later periods, because it is proportional to $\gamma_{t|t}$, which has a hump shape. The consistently higher expected inflation pushes down g and raises π. As the private-sector estimate of γ approaches the truth, the central bank approaches the classic discretionary inflation bias outcome of significant inflation and minimal output gap.

For all three setups with partial information, LI, DI, and Discretion, the policymaker chooses to weight the persistent shock less, lowering the accuracy of the private sector's information set. This effect is most measurable for the Discretionary central bank. In the Mertens (2016) algorithm, there is a first order condition that yields the optimal discretionary response to both $\gamma_{t|t-1}$ and $\varepsilon_{t|t-1}$, even thought the latter always zero. The discretionary policy can be rewritten as

$$\pi_t^{Disc} = \begin{bmatrix} 0.21 & 0.44 & 0.26 & 0.39 \end{bmatrix} \begin{bmatrix} \gamma_t - \gamma_{t|t-1} \\ \varepsilon_t - \varepsilon_{t|t-1} \\ \gamma_{t|t-1} \\ \varepsilon_{t|t-1} \end{bmatrix}$$

and we can see that less weight is placed on $\left(\gamma_t - \gamma_{t|t-1}\right)$ than $\gamma_{t|t-1}$, and more weight is placed on $\left(\varepsilon_t - \varepsilon_{t|t-1}\right)$ than $\varepsilon_{t|t-1}$. The discretionary central banker cannot affect its future actions, but it still cares about future utility, and those differences are because the weights on the predictable components of γ and ε don't affect the information of the private sector next period, but the weights on the prediction errors do. With regards to the to the prediction errors, less weight on γ and more weight on ε lowers expected inflation today, as well as improving the informational position of the next period's policymaker. Because the central bank with commitment bases its predictable inflation policy of $\pi_{t|t-1}$ on both $\gamma_{t-1|t-1}$ and π_{t-1}, it is not as easy to compare prediction errors for surprise in γ_t with those predictable behaviors. Nevertheless, we see that for both LI and DI, more weight is placed on the transient shock, which worsens the private sector predictions.

2.6 Conclusion

In this paper, I have presented three novel formulations, capable of solving optimal linear policy for a policymaker with commitment and an informational advantage over the private sector. To find optimal policy, all formulations must be based on the unconditional perspective, meaning the policymaker commits to their linear weightings before any shocks are realized. The recursive formulation enables the finding of a steady state, in which the policymaker takes significant advantage of both commitment and shaping of the private sector's expectations.

3

Intermediate Commitment, Temptation, and Central Banks

3.1 Introduction

It is well known that central banks face a time-inconsistency problem. The optimal choices from this period's perspective mitigate today's constraint via the expectations channel. From next period's perspective, however, the central bank would have lower costs by ignoring this period's constraint and shocks. In standard optimal monetary policy, central banks are divided into discretionary banks and those with commitment. Discretionary central banks reoptimize every period, which means that next period's central bank will adjust its decision to lower costs today. Those with commitment make a state-contingent plan and bind themselves in future periods to follow through upon the plan, so next period actions are adjusted to reduce costs for this period.

The standard central bank with commitment ignores the past in the very first period, but ever after follows through on a state-contingent plan in which actions are responsive to the prior period's constraint. There is some tension in one type of policy the very first period, and then quite a different one in later periods. To address this tension, Woodford (1999) and later work advocates for central bankers to follow a "timeless perspective." Such a central bank has commitment about future periods and *chooses* in the first period to follow through on what it would have committed itself to. In practice, this means every period's decision incorporating the prior period's constraint.

This paper models a central bank that, every period wishes to follow Woodford's advice, but does not manage to fully implement it. Marcet and Marimon (2019) and Giannoni and Woodford (2017) show that timeless policy is equivalent to incorporating the prior period's Lagrange multiplier into this period's utility function. The central bank with *Scaled Commitment* applies a discount factor between zero and one to prior period's Lagrange multiplier. It puts the degree of commitment on a spectrum, where a discount factor of zero, i.e. ignoring the prior constraint, is discretion and a factor of one, i.e. fully incorporating it, is commitment.

I extend a version of Recursive Contracts to accommodate intermediate forms of commitment for models with one-period forward-looking constraints. I apply this to textbook a New Keynesian monetary model and derive how a central bank with Scaled Commitment compares with discretion and commitment. In particular, it has correlation across periods, unlike discretion, and still some persistent effects of one-time shocks, unlike commitment. Both behaviors are intermediate between discretion and commitment. The key benefit of modeling a Scaled Commitment central bank is a consistent description that could yield such behavior, and the possibility of estimating the parameter in future work. Scaled Commitment also allows for different discounts to different prior-period constraints. I explore the implications if the policymaker follows through on commitment regarding the always-binding Phillips curve, but does not have commitment regarding the occasionally binding ZLB.

The most similar work is that of Loose Commitment used first in Debortoli and Nunes (2010). In it, there is a stochastic chance every period that the central bank re-optimizes, and breaks with past commitments. Both Loose and Scaled Commitment have similar effects of reducing the effectiveness of the expectations channel to mitigate current period shocks. They have different effects when the next period arrives. Scaled Commitment places a limit on the intertemporal sacrifice that next period's policymaker is willing to make to accommodate this period's shocks. Loose Commitment is uncertainty about whether the promises made today will be followed through upon. Higher probability of reoptimization can lead to more extreme promises, which does not have a comparable behavior in Scaled Commitment.

Beyond the applications, I present a general extension of Recursive Contracts that can nest more complicated discounting functions of prior Lagrange multipliers. I also show how Scaled Commitment is a simple modification to existing solution methods for Linear Quadratic Regulator models for solving commitment.

The paper proceeds as follows. Section 3.2 develops the two-period model and applies Gul and Pesendorfer (2001) temptation. Section 3.3 derives the recursive result before providing some results of scaled commitment. Section 3.4 solves intermediate commitment models in a more general context, as well as solving scaled commitment in a more general linear setup.

3.2 Time inconsistency in a two-period model

In this section, I consider a simple two-period New Keynesian monetary model. I discuss time-inconsistency and formally introduce Scaled Commitment. I then derive the Gul and Pesendorfer (2001) implied temptation function implied by the intermediate behavior. Finally, I show how the general approach can also be used to express Loose Commitment.

In a New Keynesian model, the central bank chooses its interest rates over time,

i_t. The equilibrium conditions for the private sector agents ensure two things[1]

$$y_t = -i_t + E_t\left[y_{t+1} + \pi_{t+1}\right]$$

$$\pi_t = y_t + \varepsilon_t + \beta E_t\left[\pi_{t+1}\right]$$

$$\varepsilon_t \sim N\left(0, \sigma^2\right)$$

where y_t is the output gap, π_t is the inflation, and ε_t is the cost-push shock. The first constraint is the dynamic IS equation, and the second one is the New Keynesian Phillips curve. These are both forward-looking constraints. First, we assume that there is no zero lower bound limiting i_t. Next, following the Ramsey approach, we let the central bank choose y_t and π_t subject to the Phillips curve.

Now we simplify the problem even further by assuming only two periods (and dropping the β). The constraints on the choice of πs and ys are

$$\pi_1 = y_1 + \varepsilon_1 + E_1\left[\pi_2\right] \tag{3.1}$$

$$\pi_2 = y_2 + \varepsilon_2$$

where ε_1 and ε_2 are the independent and identically distributed mean-0 shocks, with standard deviation σ. They are known each period before the corresponding π and y are chosen. The expectations operator comes from the optimality condition for the private sector agents, thus it is their expectations that matter.

Each period, the second order approximation of the social welfare is,[2]

$$U_t = -\frac{1}{2}\left(\pi_t^2 + y_t^2\right)$$

Consider the central bank in period 1 with this utility and constraint. From the bank's perspective, the ideal would be if they could manipulate the private-sector

[1] I have dropped all constants except β for simplicity.

[2] For simplicity again, I drop the constant for the relative weight on the output gap.

beliefs away from rational expectations,

$$\pi_1 = y_1 = 0$$
$$E_1[\pi_2] = -\varepsilon_1$$
$$\pi_2 = -y_2 = \frac{\varepsilon_2}{2}$$

They would not incur any costs in the first period, and the second period losses would be at the discretionary minimum. However, rational expectations explicitly disallow this. If the central bank in period 2 is going to ignore ε_1, then $E_1[\pi_2] = 0$.

The central bank will adjust both π_1 and y_1 to meet (3.1).[3] To produce the highest overall utility, they would also share the adjustment burden with the second period by (honestly) changing $E_1[\pi_2]$. However, when period 2 comes, the central bank would prefer to ignore ε_1 in their decision making.

The standard analysis divides the central banks into discretionary ones who ignore ε_1 once period 2 arrives, and those with commitment who follow through on the optimal incorporation from the period 1. We will first work through the model for both these central bank types and then explore intermediate behavior.

The Lagrangian for a discretionary bank in the second period is,

$$\mathcal{L} = -\frac{1}{2}\left(\pi_2^2 + y_2^2\right) + \gamma_2\left(\pi_2 - y_2 - \varepsilon_2\right)$$

Then

$$\pi_2 = \gamma_2 = \frac{\varepsilon_2}{2}$$
$$y_2 = -\gamma_2 = -\frac{\varepsilon_2}{2}$$

Second period unconditional expected utility is

$$E\left[-\frac{1}{2}\left(\pi_2^2 + y_2^2\right)\right] = -\frac{1}{4}\sigma^2$$

[3] Because they are equally weighted in the utility and in the constraint, they will split the adjustment across them.

Because ε_2 is mean 0, $E_1[\pi_2] = 0$. The solution is the same for the first period, and we have

$$\pi_1 = \gamma_1 = \frac{\varepsilon_1}{2}$$

$$y_1 = -\gamma_1 = -\frac{\varepsilon_1}{2}$$

Banks with commitment function as if they are choosing the state-contingent policies for π_2 and y_2 at time 1. Their problem is

$$\max_{\pi_1, y_1, \pi_2, y_2} -\frac{1}{2}\left(\pi_1^2 + y_1^2 + \pi_2^2 + y_2^2\right)$$

$$\text{s.t. } \pi_1 = y_1 + \varepsilon_1 + E_1[\pi_2]$$

$$\pi_2 = y_2 + \varepsilon_2$$

Their Lagrangian is

$$\mathcal{L} = E_1\left[-\frac{1}{2}\left(\pi_1^2 + y_1^2 + \pi_2^2 + y_2^2\right) + \gamma_1(\pi_1 - y_1 - \varepsilon_1 - \pi_2) + \gamma_2(\pi_2 - y_2 - \varepsilon_2)\right]$$

Their solution is

$$\begin{aligned}
\pi_1^c &= \gamma_1 &&= \frac{2}{5}\varepsilon_1 \\
y_1^c &= -\gamma_1 &&= -\frac{2}{5}\varepsilon_1 \\
\pi_2^c &= \gamma_2 - \gamma_1 &&= -\frac{1}{5}\varepsilon_1 + \frac{1}{2}\varepsilon_2 \\
y_2^c &= -\gamma_2 &&= -\frac{1}{5}\varepsilon_1 - \frac{1}{2}\varepsilon_2
\end{aligned}$$

Following through on the commitment lowers second period utility to

$$\begin{aligned}
E\left[-\frac{1}{2}\left(\pi_2^{c2} + y_2^{c2}\right)\right] &= -.29\sigma^2 \\
&= 1.16 * \left(-\frac{1}{4}\sigma^2\right).
\end{aligned}$$

Thus the second-period central bank using commitment is 16% worse off. But, the benefit to the first-period central bank is

$$E\left[-\frac{1}{2}\left(\pi_1^{c2}+y_1^{c2}\right)\right] = -\frac{4}{25}\sigma^2$$
$$= .64 * \left(-\frac{1}{4}\sigma^2\right).$$

The first period central bank is 36% better off. This shows the utility benefits to the first period and costs in the second period.

This is a radically simplified model, but we have used it to highlight the intertemporal tradeoff a commitment central bank achieves: a small cost to second-period utility has a larger benefit to first-period utility. Commitment allows the central bank to make beneficial intertemporal sacrifice.

3.2.1 Intermediate Bank Behavior

In this subsection, I propose a method of indexing banks based on them applying a discount in the second period to the first period Lagrange multiplier. This will nest discretion and commitment: a discretionary central bank in the second period ignores the first period constraint, and a central bank with commitment in the second period chooses policies that incorporates its effect on the first period constraint.

Marcet and Marimon (2019) and others show that if the policymaker properly incorporates the first-period Lagrange multiplier into his second-period utility, then optimal second-period policy will follow through on the first-period commitment. This is an application of Woodford's encouragement to follow timeless policy in the second period: choosing to do what the policymaker would have wished you to do based on a prior perspective.

Framed as a Recursive Contract saddlepoint problem, the period problems for a bank with commitment are

$$U_1(\varepsilon_1) = \min_{\gamma_1} \max_{\pi_1, y_1} -\frac{1}{2}\left(\pi_1^2 + y_1^2\right) + \gamma_1(\pi_1 - y_1 - \varepsilon_1) + E\{U_2(\gamma_1, \varepsilon_2)\}$$

$$U_2(\gamma_1, \varepsilon_2) = \min_{\gamma_2} \max_{\pi_2, y_2} -\frac{1}{2}\left(\pi_2^2 + y_2^2\right) - \gamma_1\pi_2 + \gamma_2(\pi_2 - y_2 - \varepsilon_2)$$

The Lagrange multiplier γ_1 enforces the constraint, because the second period utility fully internalizes its effects on the first period constraint. The first order condition for the first problem is

$$0 = \pi_1 - y_1 - \varepsilon_1 + E\left\{\frac{\partial U_2}{\partial \gamma_1}\right\}$$

and by the envelope condition,

$$\frac{\partial U_2}{\partial \gamma_1} = -\pi_2^*$$

$$E\left\{\frac{\partial U_2}{\partial \gamma_1}\right\} = -\pi_{2|1}$$

$$0 = \pi_1 - y_1 - \varepsilon_1 - \pi_{2|1}$$

As discussed above

$$\gamma_1^* = \frac{2}{5}\varepsilon_1$$

$$\gamma_2^* = \frac{1}{2}(\varepsilon_2 + \gamma_1)$$

$$\pi_2^* = \gamma_2^* - \gamma_1^*$$

$$= \frac{1}{2}\varepsilon_2 - \frac{1}{2}\gamma_1^*$$

$$\pi_{2|1}^* = -\frac{1}{2}\gamma_1^*$$

We can also calculate explicitly

$$\begin{aligned} E\left\{U_2\right\} &= -\frac{1}{2}\sigma_{\varepsilon_2}^2 + \frac{1}{4}\gamma_1^2 \\ \frac{\partial E\left\{U_2\right\}}{\partial \gamma_1} &= \frac{1}{2}\gamma_1 \\ &= -\pi_{2|1}^* \end{aligned}$$

Definition 20. A two-period central bank with *scaled commitment* τ optimizes the following value functions,

$$\begin{aligned} U_1^{\tau}\left(\varepsilon_1\right) &= \min_{\gamma_1}\max_{\pi_1, y_1} -\frac{1}{2}\left(\pi_1^2 + y_1^2\right) + \gamma_1\left(\pi_2 - y_2 - \varepsilon_2\right) \\ &\quad + U_1^{\tau,adj}\left(\gamma_1, \pi_{2|1}^*\left(\varepsilon_1\right)\right) + E\left\{U_2^{\tau}\left(\gamma_1, \varepsilon_2\right)\right\} \\ U_1^{\tau,adj}\left(\gamma_1, \pi_{2|1}^*\left(\varepsilon_1\right)\right) &= -\left(1-\tau\right)\gamma_1\pi_{2|1}^*\left(\varepsilon_1\right) \\ U_2^{\tau}\left(\gamma_1, \varepsilon_2\right) &= \min_{\gamma_2}\max_{\pi_2, y_2} -\frac{1}{2}\left(\pi_2^2 + y_2^2\right) - \tau\gamma_1\pi_2 + \gamma_2\left(\pi_2 - y_2 - \varepsilon_2\right) \end{aligned}$$

where ε_t is a cost-push shock, π_t is inflation, y_t is the output gap, γ_t is the Lagrange multiplier, and $\pi_{2|1}^*\left(\varepsilon_1\right)$ is the reduced form of $\pi_{2|1}$ in terms of the fundamental shock ε_1, not including γ_1. It must be consistent

$$\pi_{2|1}^*\left(\varepsilon_1\right) = E\left\{\pi_2 | \varepsilon_1\right\}$$

Start by considering the second period utility function. Its only difference from the Recursive Contract one is the discount facto τ applied to the $\gamma_1\pi_2$. I interpret this as in the second period, the central banker wanting to implement timeless policy, but only able to do it up a factor of τ. An alternative interpretation is that there are two groups within the central bank, one wishing to act in the discretionary manner, and one wishing to act with commitment. The coefficient τ is a reduced-form representation of the bargaining power of the commitment group. Because the second period bank's utility may incompletely incorporate there is a utility adjustment factor $U_1^{\tau,adj}$ so that the total utility for U_1^τ accurately reflects the constraint, i.e. at the optimum $U_1^\tau(\varepsilon_1) = -L_1 + E\{L_2\}$ and the constraints are met.

Proposition 21. *Two-period banks with scaled commitment meet the Phillips Curve both periods, and the U_1^τ is accurate.*

Proof. It is easy to see that in the second period a τ central bank will choose

$$\pi_2^{\tau*} = \gamma_2 - \tau\gamma_1 = \frac{1}{2}(\varepsilon_2 - \tau\gamma_1)$$

$$y_2^{\tau*} = -\gamma_2 = \frac{1}{2}(-\varepsilon_2 - \tau\gamma_1)$$

In the second period, a central bank with $\tau < 1$ discounts the benefit to the first period policymaker of his action, internalize $-\tau\gamma_1\pi_2$ instead of $-\gamma_1\pi_2$. This is precisely why it under-adjusts π_2 in response to the first period constraint. The first period utility function has the factor $U_1^{\tau,adj}(\gamma_1)$ in order to adjust its utility to accommodate the discounting by the bank in the second period. The second period's expected utility will end up being

$$E\{U_2^\tau(\gamma_1, \varepsilon_2)\} = E\{L_2\} - \tau\gamma_1\pi_{2|1}^*$$

But we want the Lagrange approach to be

$$U_2^\tau(\varepsilon_1) = L_1 + E\{L_2\} + \gamma_1(\pi_1 - y - \varepsilon_1) - \gamma_1\pi_{2|1}$$

not, $-\tau\gamma_1\pi_{2|1}^*$. Therefore, we use

$$\begin{aligned} U_1^{\tau,adj}\left(\gamma_1, \pi_{2|1}^*(\varepsilon_1)\right) &= -(1-\tau)\gamma_1\pi_{2|1}^*(\varepsilon_1) \\ &= -\gamma_1\pi_{2|1}^*(\varepsilon_1) + \tau\gamma_1\pi_{2|1}^*(\varepsilon_1) \end{aligned}$$

It is crucial that the $\pi_{2|1}^*$ inside $U_1^{\tau,adj}$ be unresponsive to changing γ_1 so that the saddle-point setup enforces the constraint,

$$\frac{\partial U_1^{\tau,adj}}{\partial \gamma_1} = -(1-\tau)\,\pi_{2|1}^*(\varepsilon_1)$$

$$\begin{aligned}\frac{\partial\left(U_1^{\tau,adj} + E\left\{U_2^{\tau}\right\}\right)}{\partial \gamma_1} &= -(1-\tau)\,\pi_{2|1}^*(\varepsilon_1) - \tau\pi_{2|1}^* \\ &= -\pi_{2|1}^*(\varepsilon_1)\end{aligned}$$

where the second equation takes advantage of the envelope condition inside U_2. Then the entire first order condition for γ_1 will be

$$0 = \pi_1 - y_1 - \varepsilon_1 - \pi_{2|1}^*(\varepsilon_1)$$

Then we can solve for equilibrium behavior,

$$\begin{aligned}-y_1 &= \pi_1 = \gamma_1 \\ 0 &= \gamma_1\left(2 + \frac{\tau}{2}\right) - \varepsilon_1\end{aligned}$$

yielding

$$\begin{aligned}\pi_1^* &= \frac{2}{4+\tau}\varepsilon_1 \\ y_1^* &= -\frac{2}{4+\tau}\varepsilon_1 \\ \pi_{2|1}^*(\varepsilon_1) &= -\frac{\tau}{4+\tau}\varepsilon_1 \\ \pi_2^* &= -\frac{\tau}{4+\tau}\varepsilon_1 + \frac{1}{2}\varepsilon_2 \\ y_2^* &= -\frac{\tau}{4+\tau}\varepsilon_1 - \frac{1}{2}\varepsilon_2\end{aligned}$$

□

We can see how Scaled Commitment nests discretion and commitment, by comparing $\tau = 0$ and $\tau = 1$ outcomes to those derived above.

3.2.2 Applying Gul and Pesendorfer (2001) to Scaled Commitment

In their paper, Gul and Pesendorfer (2004) (hereafter GP) derive a representation result for the utility of choice sets of actions. They show that if an actor has a certain type of preferences over sets, then their utility over choice sets is representable as two functions, the utility under commitment, and a temptation function. In this section, I show that Scaled Commitment meets their axioms, and therefore it is useful to interpret the reduced form Scaled Commitment behavior as responding to the temptation to break with the past.

The axioms in GP allow them to model situations when the availability an ex-ante inferior choice can lower overall utility. Let $A \succsim B$ indicate that choice set A is weakly preferred to choice set B. The novel axiom for GP agents is Set Betweenness. It states that if $A \succsim B$, then $A \succsim A \cup B \succsim B$.[4] An agent has a Preference for Commitment if there exists an A where $A \succ A \cup B$, with $\succ$ indicating strict preference.

The choice sets we will consider are for period 2 inflation, and will be labeled $W_{l,h}$, will represent the weight the second-period central bank puts on the first period shock: $w \in W_{l,h} \equiv [l, h]$ will correspond to $\pi_2(w) = -w\varepsilon_1 + \frac{1}{2}\varepsilon_2$

Phrased as a choice set problem, we can define the second period utility for type τ as

$$U_2^\tau(\gamma_1, \varepsilon_2; w) \equiv -\frac{1}{2}\left(\pi_2^2 + (\pi_2 - \varepsilon_2)^2\right) - \tau\gamma_1\pi_2$$

$$\text{s.t. } \pi_2 = -w\varepsilon_1 + \frac{1}{2}\varepsilon_2$$

Then the second period central bank will choose

$$w_{l,h}^\tau \equiv \arg\max_{w \in W_{l,h}} U_2^\tau(\gamma_1, \varepsilon_2; w)$$

and we can see that they will always choose the w closest to free choice above $\frac{\tau}{2+\tau}$

[4] Maximizers without time-inconsistency would be indifferent ($\sim$), $A \succsim B \implies A \sim A \cup B \succsim B$.

from $\pi_2^\tau = -\frac{\tau}{2+\tau}\varepsilon_1 + \frac{1}{2}\varepsilon_2$,

$$w_{l,h}^\tau = \arg \min_{w \in \left[l,\, h\right]} \left| w - \frac{\tau}{2+\tau} \right|$$

We can define $\pi_2^{\tau,l,h}$ as the inflation chosen by type τ from the choice set $W_{l,h}$, and $\pi_{2|1}^{\tau,l,h}$ its first period expectation,

$$\pi_2^{\tau,l,h} = -w_{l,h}^\tau \varepsilon_1 + \frac{1}{2}\varepsilon_2$$

$$\pi_{2|1}^{\tau,l,h} = -w_{l,h}^\tau \varepsilon_1$$

therefore,

$$E\left\{L_2 | w_{l,h}^\tau, \varepsilon_1\right\} = -\frac{1}{4}\sigma_{\varepsilon_2} - \left(w_{l,h}^\tau\right)^2 \varepsilon_1^2$$

Given $w_{l,h}^\tau$, we can also solve for the period 1 choices

$$\gamma_1 = \pi_1^{\tau,l,h} = -y_1^{\tau,l,h} = \frac{1}{2}\left(1 - w_{l,h}^\tau\right)\varepsilon_1$$

and utility from the first period perspective

$$U_1^\tau\left(\varepsilon_1; w_{l,h}^\tau\right) = -\left(\frac{1}{4}\left(1 - w_{l,h}^\tau\right)^2 + \left(w_{l,h}^\tau\right)^2\right)\varepsilon_1^2 - \frac{1}{4}\sigma_{\varepsilon_2}^2$$

This utility is a quadratic function of $w_{l,h}^\tau$ with a minimum at 1/5. Therefore, it exhibits Set Betweenness, with $\succsim_\tau$ representing the preferences of a central bank with Scaled Commitment τ,

$$W_{l,h} \succsim_\tau W_{l',h'} \implies \left|w_{l,h}^\tau - \frac{1}{5}\right| \leq \left|w_{\min(l,l'),\max(h,h')}^\tau - \frac{1}{5}\right| \leq \left|w_{l',h'}^\tau - \frac{1}{5}\right|$$

$$\iff W_{l,h} \succsim_\tau \left(W_{l,h} \cup W_{l',h'}\right) \succsim_\tau W_{l',h'}$$

The model described meets the other axioms of the GP result: they are complete, they satisfy continuity, and they satisfy the independence axiom. Thus, their representation result holds

$$V_1^\tau\left(W_{l,h}; \varepsilon_1\right) = \max_{w \in W_{l,h}} u\left(w; \varepsilon_1, \tau\right) + v\left(w; \varepsilon_1, \tau\right) - \max_{\tilde{w} \in W_{l,h}} v\left(\tilde{w}; \varepsilon_1, \tau\right) \tag{3.2}$$

Proposition 22. *The following definitions of u and v represent a two-period central bank with intermediate commitment $\tau \in (0,1]$ over choice sets $W_{l,h}$ in equation (3.2), for shock ε_1*

$$u(w;\varepsilon_1,\tau) = -\left(\frac{1}{4}(1-w)^2 + w^2\right)\varepsilon_1^2$$

$$v(w;\varepsilon_1,\tau) = -\left(\frac{1}{\tau} - 1\right) w^2 \varepsilon_1^2$$

Proof. Inside $V_1^\tau(W_{l,h};\varepsilon_1)$, the proof amounts to checking that the function u represents the utility of singleton menus; and w^* represents the value actually chosen.

Singleton menus $W_{w,w} = \{w\}$ force the choice,

$$\pi_2^{\tau,w,w} = -w\varepsilon_1 + \frac{1}{2}\varepsilon_2$$

and this has the same relative utility of for different w as U_1^τ, therefore it is identified up to affine transforms, and I omit the constant in U_1^τ.

Now let's check that $w^* = w_{l,h}^\tau$ from above.

$$u(w;\varepsilon_1,\tau) + v(w;\varepsilon_1,\tau) = -\left(\frac{1}{4}(1-w)^2 + \frac{1}{\tau}w^2\right)\varepsilon_1^2$$

It has an optimum at $\frac{\tau}{4+\tau}$, so,

$$w_{l,h}^\tau = \arg\max_{w\in W_{l,h}} u(w;\varepsilon_1,\tau) + v(w;\varepsilon_1,\tau)$$

as desired. □

The fact that GP applies supports the interpretation of time-inconsistency as a temptation. Always, the central bank at time 2 would prefer to ignore the past constraint. In the GP notation above, the most tempting element for v is always $w = 0$. From the perspective of of period 1, the central bank has a preference for commitment. I claim that the central bank in time 2 knows they should incorporate the prior constraint, and follow Woodford's advice to act in a timeless manner. Scaled Commitment allows us to put their ability to follow the timeless perspective on a spectrum. Rational expectations means that ability feeds back into first period expectations and we solve the model accordingly.

To summarize, we defined central banks with scaled commitment τ as those in the second-period decision making discount the first-period Lagrange multiplier by τ. We derived first and second period utility extensions of Recursive Contracts that were consistent with the second-period decision making. Finally, we showed that this behavior and preferences had Set Betweenness from Gul and Pesendorfer (2001). We derived u and v functions for their representation result.

3.2.3 Rephrasing Loose Commitment

The approach of a U_2 that doesn't fully match Recursive Contracts paired with a U_1^{adj} to accommodate the deficit can be used to represent Loose Commitment as well. In the previous subsection, we modeled second period policymaker as discounting the prior Lagrange multiplier by τ. That yielded $\pi_2 = \gamma_2 - \tau\gamma_1$. With respect to the index τ, the second period outcome smoothly depended upon first period constraint. In general if $\pi_2(\gamma_1, \gamma_2)$ is specified, and $E_1\left[\frac{\partial \pi_2}{\partial \gamma_1}\right] \leq 0$, then one can solve the model. The specific values of γ_1 and γ_2 will of course have to adjust so that the constraints hold, but the Lagrangian will still work.

Now I apply the same approach to Loose Commitment, (Debortoli and Nunes (2010); Bodenstein et al. (2012); Debortoli and Nunes (2014); Debortoli and Lakdawala (2016)). In this framework, the policymaker has a stochastic chance to reoptimize in the second period. If he does reoptimize, he ignores the past and π_2 does not depend on γ_1: $\pi_2 = \gamma_2^R$. If he does not reoptimize, then he follows through on the plan it made in previous period: $\pi_2 = \gamma_2^C - \gamma_1$. Note that the Lagrange multiplier under reoptimization, γ^R, will differ from the one under following through on commitment, γ^C. To describe this in an extension of Recursive Contracts, we can just add an independent, exogenous Bernoulli random variable with probability p, C. When $C = 1$, that indicates the policymaker follows through on the commitment, $\pi_2 = \gamma_2 - C\gamma_1$.[5]

Definition 23. A two-period central bank with *loose commitment probability* p optimizes the following value functions,

$$U_1^{lc,p}(\varepsilon_1) = \min_{\gamma_1} \max_{\pi_1, y_1} -\frac{1}{2}\left(\pi_1^2 + y_1^2\right) + \gamma_1\left(\pi_2 - y_2 - \varepsilon_2\right)$$

$$+ U_1^{lc,p,adj} + E\left\{U_2^{lc,p}\left(\gamma_1, \varepsilon_2, C\right)\right\}$$

$$U_1^{lc,p,adj} = (1-p)\, E\left\{\pi_2^*\left(\varepsilon_1, \varepsilon_2, C\right)|C=0\right\}$$

$$U_2^{lc,p}\left(\gamma_1, \varepsilon_2, C\right) = \min_{\gamma_2} \max_{\pi_2, y_2} -\frac{1}{2}\left(\pi_2^2 + y_2^2\right) - C\gamma_1\pi_2 + \gamma_2\left(\pi_2 - y_2 - \varepsilon_2\right)$$

where ε_t is a cost-push shock, π_t is inflation, y_t is the output gap, γ_t is the Lagrange multiplier, C is a Bernoulli random variable of probability p representing the chance the central bank does not reoptimize, and π_2^* is the equilibrium policy for π_2^*.

Solving the system for the second period policy

$$\pi_2^* = \begin{cases} \frac{1}{2}\varepsilon_2 & C = 0 \\ \frac{1}{2}\left(\varepsilon_2 - \gamma_1\right) & C = 1 \end{cases}$$

therefore, because $\varepsilon_{2|1} = 0$, $U_1^{lc,p,adj} = 0$, no adjustment is needed to the first period utility.

$$E\left\{U_2^{lc,p}\left(\gamma_1, \varepsilon_2, C\right)\right\} = E\left\{L_2\right\} - pE\left\{\gamma_1\pi_2|C=1\right\}$$

$$= E\left\{L_2\right\} - \gamma_1\pi_{2|1}$$

[5] Dropping the superscripts, just note that γ_2 depends on γ_1, ε_2, and R .

which is precisely what's needed.

Solving the system yields an equilibrium condition for the first period Lagrange multiplier:

$$4\gamma_1 + E\left[C\gamma_1\right] = 2\varepsilon_1.$$

That yields the result

$$\begin{aligned}
\pi_1^* &= \gamma_1 && = \frac{2}{4+p}\varepsilon_1 \\
y_1^* &= -\gamma_1 && = -\frac{2}{4+p}\varepsilon_1 \\
\pi_2^* &= \gamma_2 - C\gamma_1 && = -\frac{C}{4+p}\varepsilon_1 + \frac{1}{2}\varepsilon_2 \\
y_2^* &= -\gamma_2 && = -\frac{C}{4+p}\varepsilon_1 - \frac{1}{2}\varepsilon_2
\end{aligned}$$

The outcomes in the first period are quite similar to Scaled Commitment. Here the chance of not reoptimizing, p, maps to the discount τ. Both vary from 0 to 1, with $p = 0$ nesting discretion, and $p = 1$ nesting commitment. The difference appears in period 2. If the reoptimization does occur, $C = 0$, and ε_1 has no effect on the second period outcomes. If reoptimization does not occur, $C = 0$, and the second period variables react more strongly to ε_1 than if they had been under commitment $(1/\left(4+p\right) > 1/5)$.

The two applications in this section show that extending Recursive Contracts to accommodate intermediate forms commitment is a promising avenue of analysis. In the first, I defined a two-period central bank with Scaled Commitment. I showed how the value functions enforce the constraint, and how they exhibit Set Betweenness from Gul and Pesendorfer (2004). This fact supports the interpretation of the second-period central bank producing intermediate behavior in response to the temptation to break with the past. In the second exercise, I used the same type of Recursive Contracts extension to model two-period Loose Commitment. This is quite similar to the solution method used in Bodenstein et al. (2012), however, I have consolidated the value functions into one function. I provide the alternative interpretation that an intermediately committed central bank responds to shock C, not C determines which central bank shows up next period.

3.3 Recursive Intermediate Commitment

The previous section focuses on the two-period central bank model. This section extends the reduced-form Lagrange discounting into a recursive framework. We also reintroduce standard constants (κ and λ) and allow a positive output gap.

$$\pi_t = \kappa y_t + \varepsilon_t + \beta E_t \left[\pi_{t+1}\right]$$

$$U_t = -\frac{1}{2}\left(\pi_t^2 + \lambda \left(\overline{y} - y_t\right)^2\right).$$

When considering his action this period, the central bank would prefer not to incorporate how his expected action this period affected the prior period's constraint. But he would prefer that the next period's central bank take into consideration this period's constraint.

As discussed above, the standard analysis divides central banks into discretionary ones and those with commitment. In every period t, the discretionary actor ignores the past, and makes his choices. He rightly expects the future version of himself to do the same. His Lagrangian looks like

$$\mathcal{L}^D = U + \gamma\left(\pi - \kappa y - \varepsilon - \beta E\left[\pi_{+1}\right]\right)$$

For a discretionary central banker facing the New Keynesian Phillips curve, future utility is not a consideration, because this period's actions cannot affect it. Inflation next period, π_{+1}, must be taken as given. The optimal choice here is $\pi^* = \gamma$.

In the case of the committed actor, we assume there is a date t_0 when the central bank makes his state-contingent plans. His Lagrangian looks like

$$\mathcal{L}^C = E_{t_0} \sum_{t=t_0}^{\infty} \beta^{t-t_0} \left(U_t + \gamma_t \left(\pi_t - \kappa y_t - \varepsilon_t - \beta \pi_{t+1}\right)\right) \tag{3.3}$$

This problem produces one policy for t_0, and a different one for all $t > t_0$: $\pi^*_{t_0} = \gamma_{t_0}$; and for all $t > t_0$, $\pi^*_t = \gamma_t - \gamma_{t-1}$. In the first period, the central bank ignores the past, and the forever after incorporates it.

The discrepancy between the first and subsequent periods led Woodford (1999) to advocate the central bank follow "a pattern of behavior *to which it would have wished to commit itself to at a date far in the past*." As he expanded in Woodford (2003), "Rather than doing one thing now but promising to behave differently in the future, one should follow a time-invariant policy that is of the kind that one would always wish *to have been expected* to follow." Timeless monetary policy argues the central bank choose $\pi^*_t = \gamma_t - \gamma_{t-1}$ for the first period as well as all subsequent periods.

Central banks have some control over how they choose their policy, but they may not be impervious to the temptation to ignore the past. I derive a method to answer the question, "what if central banks imperfectly follow Woodford's advice?" I express this behavior recursively. If the central bank has commitment, Marcet and Marimon (2019) show how to use Recursive Contracts to solve the problem as a value function. They transform it into a saddle-point problem by adding the previous period's Lagrange multiplier as a state variable to the value function and the decision making.

$$V^C\left(\varepsilon, \gamma_{-1}\right) = \min_{\gamma} \max_{\pi, y} U + \gamma\left(\pi - \kappa y - \varepsilon\right) - \gamma_{-1}\pi + \beta E\left[V\left(\varepsilon', \gamma\right)\right] \tag{3.4}$$

where superscript C is for commitment, ε is the cost-push shock, γ_{-1} is the prior period's Lagrange multiplier, γ is this period's Lagrange multiplier, π is this period's inflation, y is this period's output gap, and $\beta \in [0, 1)$. For the central banker in this model, the timeless choice is optimal, and the FOC for γ, along with the envelope condition, ensure that the constraint holds.

Recursive Contracts are often interpreted as a numerical approach to the true underlying Lagrangian. In that interpretation, the true problem is represented by (3.3), and (3.4) is this the economic modeler's method of solving for the policymaker's plan. However, I suggest that the recursive framing is actually a better representation of the problem facing the central banker. There is no common knowledge, state-contingent policy written down ready to be followed through upon. Woodford was speaking to the central bankers *today*, and encouraging them to internalize their choices' effects on prior constraints. How much should they internalize them? $\gamma_{-1}\pi$. In Giannoni and Woodford (2017), they describe how incorporating the prior period's Lagrange multiplier into this period's social welfare consideration is an alternate route to timeless policy. The fact that Woodford and others encourage policymakers to act timelessly highlights that there actually are policymakers deciding at every time t. If they operate in a timeless perspective, they incorporate $\gamma_{-1}\pi$.

My generalization substitutes an arbitrary function $\mathcal{T}$ in place of $\gamma_{-1}\pi$. This function represents the degree to which the central banker incorporates the prior constraint into his decision process this period,

$$V\left(u,\gamma_{-1}\right)=\min_{\gamma}\max_{\pi,y}U-\mathcal{T}\left(u,\gamma_{-1},\pi\right)+\gamma\left(\pi-\kappa y-u\right)+V^{\mathcal{T},adj}\left(u,\gamma_{-1},\gamma\right)+\beta E\left[V\left(u',\gamma\right)\right]$$

$\pi^{*}=\gamma-\mathcal{T}_{\pi}$. $V^{\mathcal{T},adj}$ is a utility adjustment in case $E\left[\mathcal{T}\left(u',\gamma,\pi'\right)\right]$ does not equal $E\left[\gamma\pi'\right]$. The Lagrange method requires two things. First,

$$V^{T,adj}\left(u,\gamma_{-1},\gamma\right)+\beta E\left[V\left(u',\gamma\right)\right]=-\gamma\beta E\left[\pi'\right]$$

so that γ, in expectation, inside $V\left(\cdot\right)$, there is a net term $\gamma\left(\pi-\kappa y-u-E\left[\pi'\right]\right)$, and when the constraint is met that term goes to 0. Second, we require that

$$\frac{\partial\left(V^{T,adj}\left(u,\gamma_{-1},\gamma\right)+\beta E\left[V\left(u',\gamma\right)\right]\right)}{\partial\gamma}=-\beta E\left[\pi'\right]$$

so that minimization of γ enforces the constraint. To do this, $V^{\mathcal{T},adj}\left(u,\gamma_{-1},\gamma\right)$ will have to depend on what $\mathcal{T}$ is used, and probably used an equilibrium formula for optimal $E\left[\pi'\right]$. If $\mathcal{T}=\gamma_{-1}\pi$, $\pi^{*}=\gamma-\gamma_{-1}$, $V^{T,adj}=0$, we get back Recursive Contracts.

Suppose the central banker reads Woodford and knows he *should* act as he would want from a prior period, but he cannot bring himself to sacrifice that much. In the language of the previous section, he knows what he should do, but he faces a temptation to ignore the past. If the temptation is overwhelming, he will act with discretion and $\mathcal{T}=0$, but it could be some intermediate level. An intermediate temptation would be represented by a $\mathcal{T}$ between 0 and $\gamma_{-1}\pi$.

Loose Commitment

As discussed in the two-period setup, the closest similar work is Loose Commitment, primarily developed by Debortoli and Nunes (Debortoli and Nunes (2010); Bodenstein et al. (2012); Debortoli and Nunes (2014); Debortoli and Lakdawala (2016)). In their framework, there is a stochastic chance of reoptimization every period. They solve for a period of reoptimization by plugging in $\gamma_{-1}=0$.[6] In their notation, they have extra terms representing the forward-looking chance of reoptimization.[7] Like in the two-period setup, we can define

$$\begin{aligned}V^{lc}\left(u,C,\gamma_{-1}\right)=\min_{\gamma}\max_{\pi,y}U-C\gamma_{-1}\pi+\gamma\left(\pi-\kappa y-u\right)\\+V^{lc,adj}\left(u,C,\gamma_{-1},\gamma\right)+\beta E\left[V^{lc}\left(u,C,\gamma_{-1}\right)\right]\end{aligned}$$

[6] Page 130 of Bodenstein et al. (2012): The value function for a period when the central bank reoptimizes is is equal to the full one with prior obligations set to zero, "$V^{R}\left(u_t,g_t\right)=V\left(u_t,g_t,\mu_t^1=0,\mu_t^2=0\right)$."

[7] For example, $h_t^{PC}\equiv\pi_t-\kappa y_t-\beta\left(1-\eta\right)E_t\pi_{t+1}^{R}-u_t$, with η being the chance of not reoptimizing.

with C an independent, exogenous Bernoulli random variable with probability p, and in equilibrium require that

$$V^{lc,adj}\left(u, C, \gamma_{-1}, \gamma\right) = -\gamma \left(1 - p\right) \beta E\left[\pi' | C' = 0\right]$$

$$\frac{\partial V^{lc,adj}\left(u, C, \gamma_{-1}, \gamma\right)}{\partial \gamma} = \left(1 - p\right) \beta E\left[\pi' | C' = 0\right]$$

which means that the value $E\left[\pi' | C = 0\right]$ must be computed from (u, C, γ_{-1}), not depending on γ.

3.3.1 Scaled Commitment

Now we will give the specific formulas for Scaled Commitment in which the Lagrangian from the prior period is discounted by τ. The central banker knows he should internalize the costs on the prior constraint, but he discounts them,

$$\mathcal{T}\left(u, \gamma_{-1}, \pi\right) = \tau \gamma_{-1} \pi$$

for $\tau \in [0, 1]$. That yields the following results

$$\pi^* = \gamma - \tau \gamma_{-1}$$

$$V^{\tau,adj}\left(u, \gamma_{-1}, \gamma\right) = -\beta \left(1 - \tau\right) \gamma E\left[\pi'^*\right]$$

where τ represents the type of central banker or the central banker's decision-making process. Note, an interior value for τ is not about breaking promises. Instead, τ controls how much the central banker is capable of spreading sacrifice across time. The Phillips curve must be met, and the question is how to spread the burden of meeting it between this period (π_t and y_t) and next period (π_{t+1}, and by implication from the $t+1$ Phillips curve, y_{t+1}). In the timeless perspective, the marginal utility of adjustments to the constraint is 1:1.[8] In Scaled Commitment, the marginal utility of the ratio of adjustments is $1 : \tau$. Whatever the previous period's marginal disutility

[8] The central banker's desire to have $1 : \beta^{-1}$ is offset by the β discount on the effectiveness of changes in the next period's decision.

for adjusting π, this period's central bank is willing to match that disutility discounted by τ. The next period's central banker will do the same with regards to this period's adjustment. This approach is akin to other forms of short-term thinking such as β-δ discounting from Laibson (1997).

The model can be solved analytically, and the difference equation becomes

$$\beta E\left[\gamma_{+1}\right] - \left(1 + \frac{\kappa^2}{\lambda} + \beta\tau\right)\gamma + \tau\gamma_{-1} = -\kappa\overline{y} - u.$$

In the context of this model, Woodford (2003) discusses discretion, t_0 optimal, and timeless policy. As scaled commitment nests these, I reproduce some of the key graphs from his discussion to illustrate the effect of τ. The results are unsurprising, yielding more or less a convex set of behaviors between commitment and discretion. Most promising is that because the decision is linear in the Lagrange multipliers, it can be translated into a LQR problem, as I describe in subsection 3.4.2.

Figure 3.1 shows the two roots of the difference equation for $\tau \in [0, 1]$. Both roots monotonically increasing, with the smaller one starting at 0, for $\tau = 0$, representing discretion.

Figure 3.2 shows the inflation bias versus τ. It is almost linear with this calibration because $\kappa^2 \gg \gamma\left(1 - \beta\right)$.

$$E\left[\pi_t\right] = \frac{\kappa\lambda\left(1 - \tau\right)}{\kappa^2 + \lambda\left(1 - \tau\right)\left(1 - \beta\right)}\overline{y}$$

I include Figure 3.3 for completeness. It shows what a t_0-optimal path would be under various settings of τ. I believe it much better to model central banks as following the same strategy across time, as opposed to the special t_0 effects a Lagrangian normally yields. Central bankers following a timeless strategy with parameter τ would produce the horizontal asymptotes that the inflation approaches.

Finally, Figure 3.4 is the impulse response to a $\varepsilon = 1$ shock at time 0. We can see that the greater the distortion from timelessness, the more significant the change

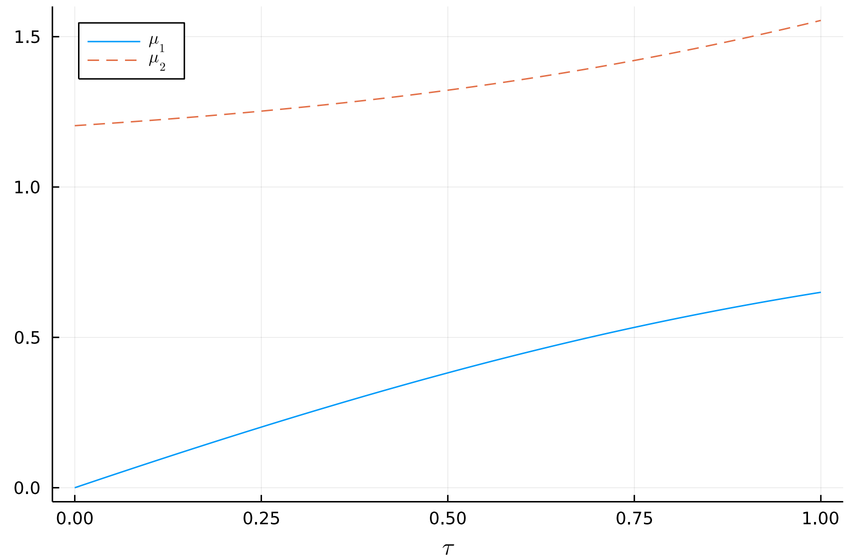

FIGURE 3.1: Difference Equation Roots

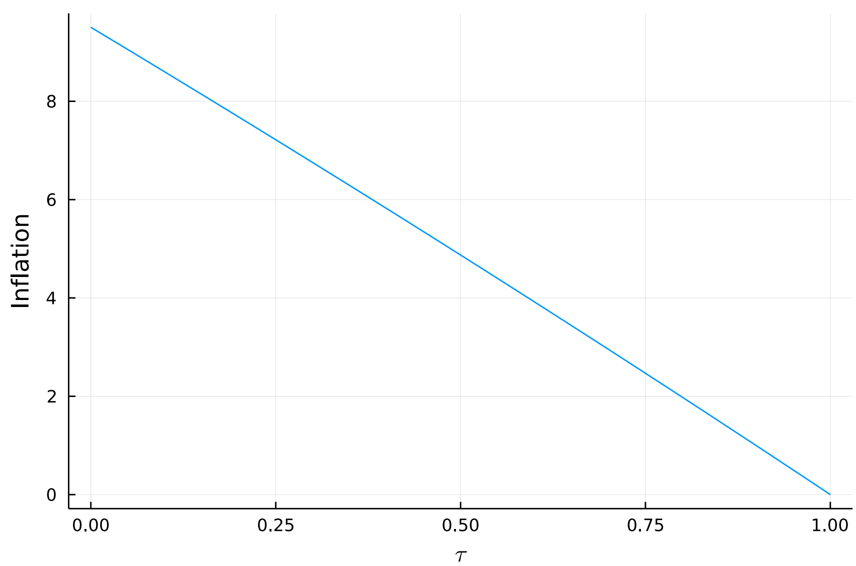

FIGURE 3.2: Annualized Inflation Bias

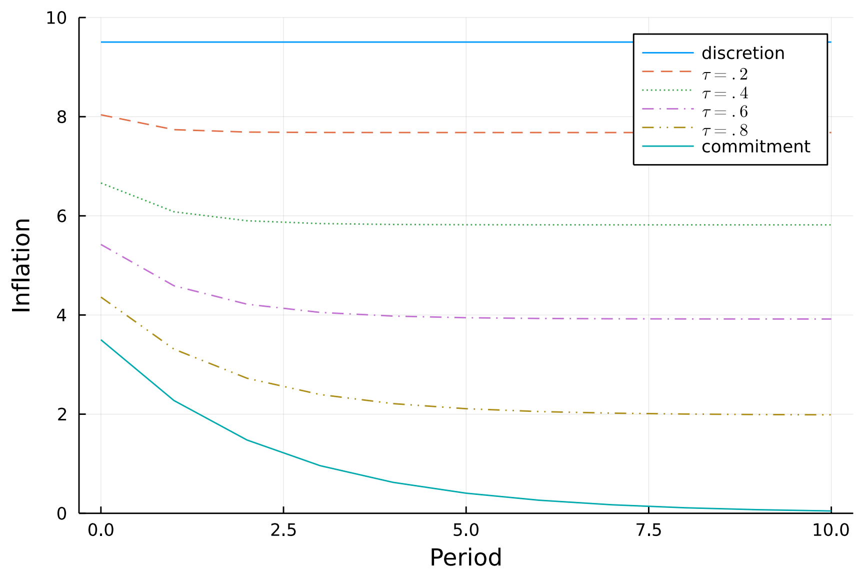

FIGURE 3.3: Perfect Foresight Equilibrium Inflation for Time-Zero Optimal Path

in output. Commitment optimally uses below expected inflation in later periods to spread the disutility across time.

3.3.2 Scaled Commitment and ZLB

Adam and Billi (2006) use Recursive Contracts to solve for optimal monetary policy incorporating the zero lower bound. In the previous subsection, we ignored the dynamic IS equation that specified how interest rates are connected to the output gap:

$$y_t = -\sigma \left(i_t - E_t \pi_{t+1}\right) + g_t + E_t \left[y_{t+1}\right].$$

The central bank could set i_t as needed to achieve any π_t and y_t that was consistent with the Phillips curve. As evident due to the Great Recession and the economic turbulence during Covid, it is also important to analyze situations where the central bank cannot lower i_t to its desired level. Adam and Billi write it as, $i_t \geq -r^*$. We can substitute this in, to phrase the constraint on y_t:

$$y_t \leq \sigma r^* + g_t + E_t \left[\sigma \pi_{t+1} + y_{t+1}\right].$$

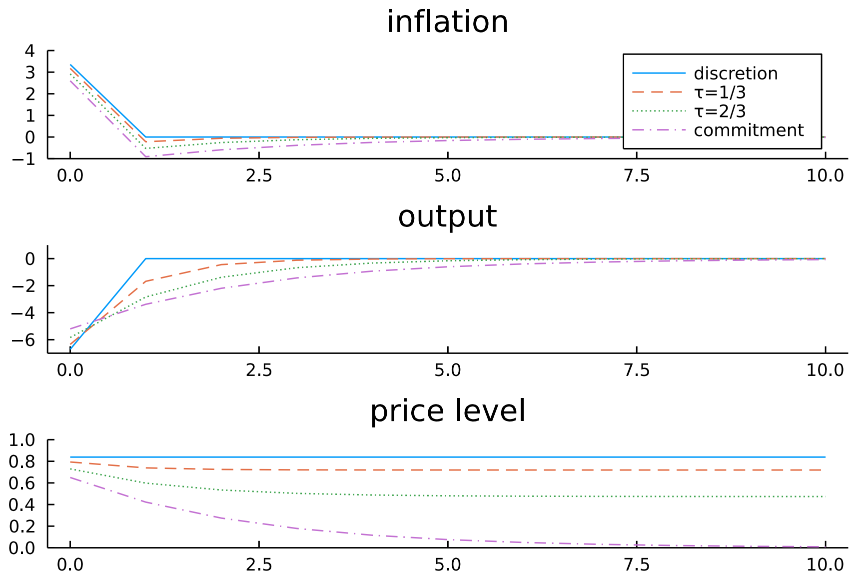

FIGURE 3.4: Impulse Response to a Temporary $\varepsilon = 1$ Cost-Push Shock

This constraint is only occasionally binding. When the economy is away from the lower bound, desired output y_t is below what would be attainable with $i_t = -r^*$. In that case, the Lagrange multiplier will be zero, $\gamma^z = 0$. In our notation it would be

$$V\left(\varepsilon, g, \gamma^p_{-1}, \gamma^z_{-1}\right) = \min_{\gamma^p \in \mathbb{R}, \gamma^z \in \mathbb{R}^+} \max_{\pi, y} \Big\{ U + \gamma^p \left(\pi - \kappa y - \varepsilon\right) + \gamma^z \left(-y + \sigma r^* + g\right)$$
$$- \mathcal{T} + V^{\mathcal{T},adj} + \beta E\left[V\left(\varepsilon', g', \gamma^p, \gamma^z\right)\right] \Big\}$$

where γ_1 is minimized over all γ^p is minimized over all of $\mathbb{R}$, because the Phillips curve always binds, and only $\gamma^z \geq 0$ is considered because the zero lower bound (ZLB) is only occasionally binding. They solve it using the timeless value $\mathcal{T} = \gamma^p_{-1}\pi - \gamma^z_{-1}\beta^{-1}\left(\sigma\pi + y\right)$.

Scaled commitment is represented by

$$\mathcal{T}\left(\pi, y, \gamma^p_{-1}, \gamma^z_{-1}\right) = \tau_p \gamma^p_{-1}\pi - \tau_z \gamma^z_{-1}\beta^{-1}\left(\sigma\pi + y\right)$$
$$V^{\mathcal{T},adj}\left(\varepsilon, g, \gamma^p_{-1}, \gamma^z_{-1}, \gamma^p, \gamma^z\right) = -\beta\left(1 - \tau_p\right)\gamma^p E\left[\pi'^*\right] + \left(1 - \tau_z\right)\gamma^z E\left[\sigma\pi'^* + y'^*\right]$$

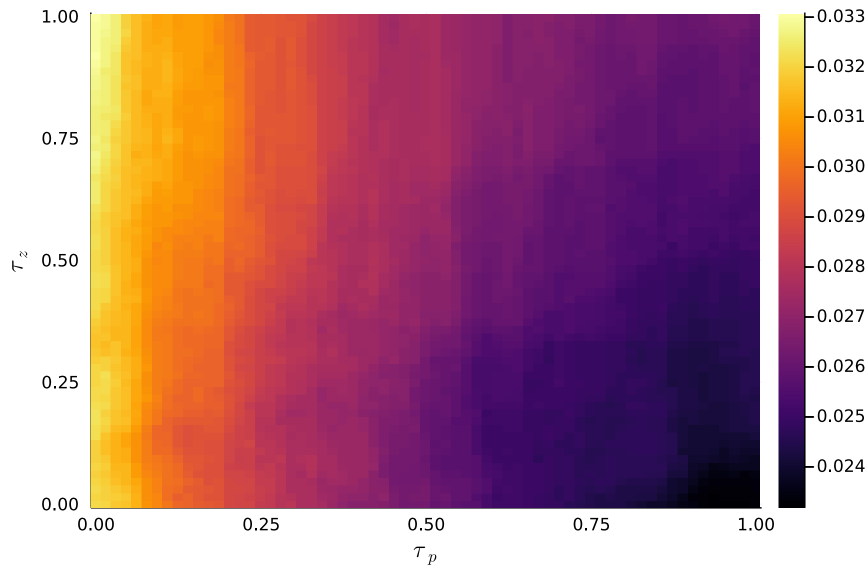

FIGURE 3.5: Risk Each Period of Reaching the Zero Lower Bound

which yields a policy of

$$\pi^* = \gamma^p - \tau_p \gamma^p_{-1} + \tau_z \gamma^z_{-1} \sigma \beta^{-1}$$

$$y^* = \frac{1}{\lambda} \left(-\gamma^p \kappa - \gamma^z + \tau_z \gamma^z_{-1} \beta^{-1} \right).$$

Figure 3.5 shows the period risk of reaching the ZLB. On the x-axis is τ_p, the measure of how much the previous period's Phillips curve is incorporated. On the y-axis is τ_z, the measure of how much the previous period's ZLB constraint is incorporated into this period's constraint. The bottom left corner represents discretion, and the top right represents commitment. Figure 3.6 shows the unconditional welfare of the parameter range. As the figure demonstrates, τ_p is much more important for the purposes of unconditional welfare than τ_z. It is because the Phillips curve always binds, whereas the ZLB only does so occasionally.

After a $-3\sigma_g$ demand shock, Figure 3.7 shows the welfare cost and Figure 3.8 shows the additional periods at the ZLB. Finally, Figure 3.9 shows the generalized impulse response after the shock. In it, τ_p represents the bottom right corner of the heatmaps, with $\tau_p = 1$ and $\tau_z = 0$. τ_z represents the converse with $\tau_z = 1$ and $\tau_p = 0$.

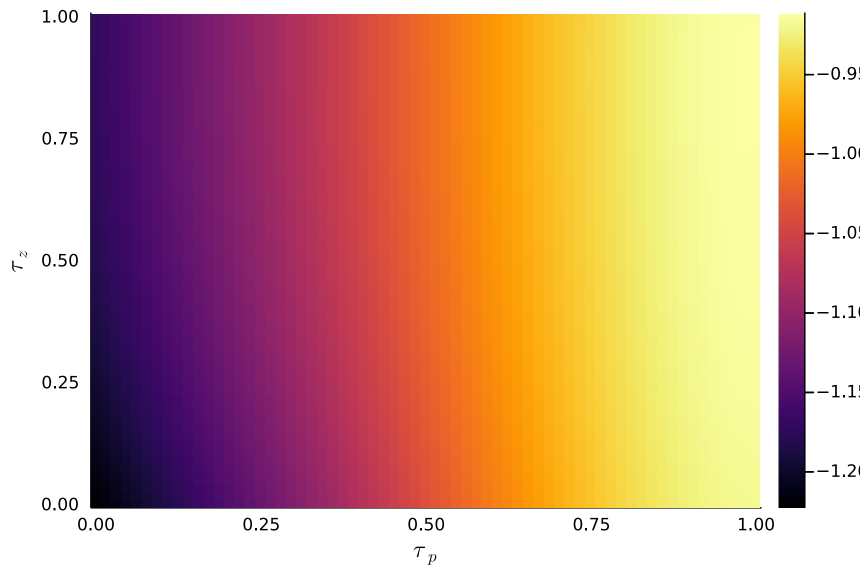

FIGURE 3.6: Expected Welfare

We can see that when it comes to this shock, commitment and τ_z are remarkably similar. However, from the welfare graph, we can see that although τ_p has a very different path from commitment and τ_z, it mitigates most of the welfare cost.

These dynamics seem to me to describe Japan for the past two decades and most of the industrial world since the Great Recession. They all had low and stable inflation going into things, significant difficulty dealing with the zero lower bound, and none of the overshooting after leaving the zero lower bound that would have been optimal. The central bankers did not promise to overshoot, actors in the market did not expect an overshoot, and we did not observe one after the fact. Note that models of a discretionary central bank incapable of overshooting after the zero lower bound must omit the inflation bias that was prevalent in analysis of the past.

Further work should investigate rationalizing τ_p. Consider a model where central banks imperfectly observe the scaled timeless parameters τ. During a period of disinflation, the central bank can increase private sector's estimation of τ_p. This would "anchor" inflation expectations at a level closer to commitment and also explain why expectations had a systematic error for a time. Full information models struggle to

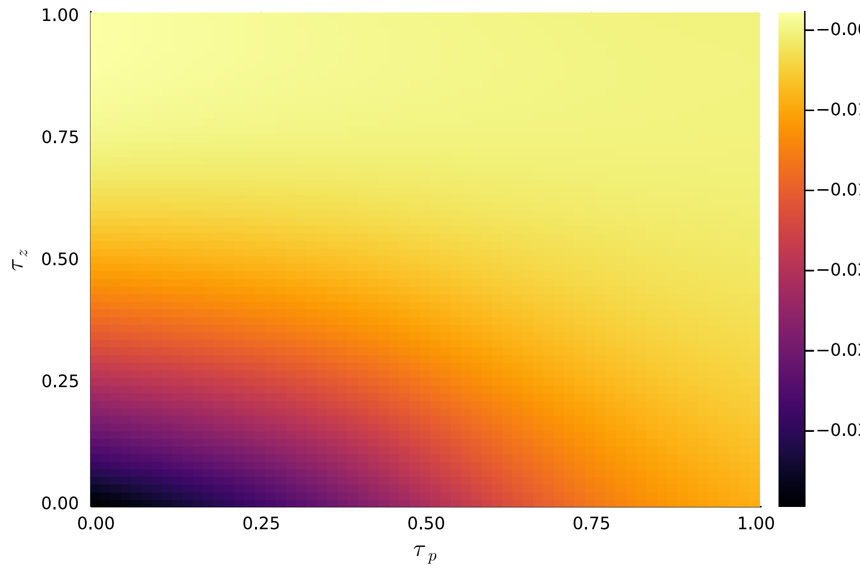

FIGURE 3.7: Welfare Cost of $-3\sigma_g$ Shock

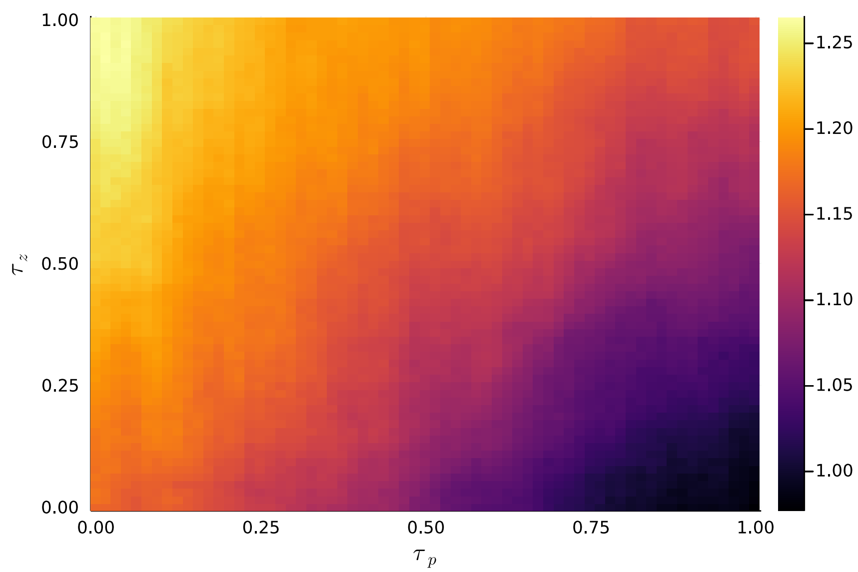

FIGURE 3.8: Additional Periods at ZLB After $-3\sigma_g$ Shock

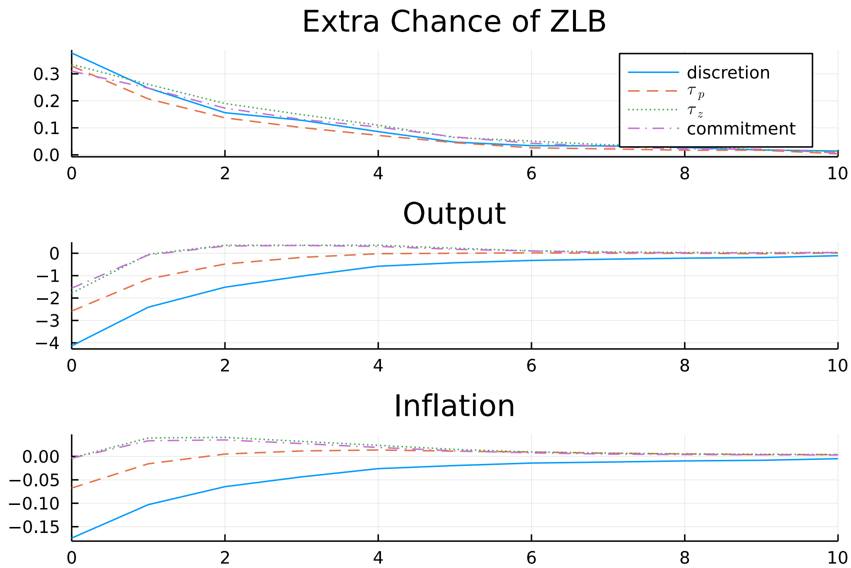

FIGURE 3.9: Impulse Response After $-3\sigma_g$ Shock

explain the consistent overestimation of expected inflation during the disinflation of the 1980s.

If the central bank only updates the private sector's estimate by acting, there are different results for the two τ's. Over time, the central bank could increase the private sector's estimate of τ_p, because every period the central bank gets the opportunity to prove its resistance to the inflation bias. However, before the first lower bound episode, there would be no way to increase the estimate of τ_z. Further, excess inflation and output after the zero lower bound would increase the perceived τ_z, but decrease the perceived τ_p. This would present the central bank with a tradeoff, and whether it was worthwhile would depend on the anticipated duration and frequency of episodes at the zero lower bound. The tradeoff would help to explain the central bank's concern that inflation expectations remain "anchored," and the frequent insistence that that inflation would not end up significantly above target, despite massive quantitative easing.[9]

[9] Beckworth (2017) shows evidence that the Federal Reserve's Quantitative Easing programs were always intended to be temporary and were never intended to drive inflation above target at a later

3.4 General Derivation

In the previous section, I worked with a simple New Keynesian problem facing a central bank. In this section, I generalize the to any problem with a one-period forward-looking constraint. This would be useful in several situations, for example, fiscal policy or a lending decision where there is a participation constraint.

We start with an infinite-time Bellman setup

$$V(x) = \max_{a \in \Gamma(x)} F(x,a) + \beta E\left[V(x') \,|x,a\right]$$

$$\text{s.t. } x' = T(x,a),$$

where $\Gamma(x)$ is the choice set, $F(x,a)$ is the instantaneous utility, and $T(x,a)$ is the transition equation, which can be stochastic. As currently set up, the model cannot handle forward-looking constraints, such as $a = x + E[a']$. Marcet and Marimon (2019) Recursive Contracts allow formulation of a time-0 or timeless agent with forward looking constraints.

$$h_0(x,a) + \beta E_t\left[h_1(x',a')\right] \geq 0,$$

where h_0 and h_1 have dimension l to handle more than one constraint. Then the problem can be reformulated as a saddle point

$$V(x,\gamma_{-1}) = \min_{\gamma \in R_+^{1\times l}} \max_a F(x,a) + \gamma h_0(x,a) + \gamma_{-1} h_1(x,a) + \beta E\left[V(x',\gamma) \,|x,a\right]$$

Note that the first-order conditions become

$$\frac{\partial F}{\partial a} + \gamma \frac{\partial h_0}{\partial a} + \gamma_{-1} \frac{\partial h_1}{\partial a} + \beta E\left[\frac{\partial V}{\partial x}\frac{\partial x'}{\partial a}\right] = 0$$

date.

and

$$\gamma_j = 0,$$

$$\text{or } h_0^j + \beta E\left[\frac{\partial V}{\partial \gamma_{-1}^j}\right] = 0$$

$$h_0^j + \beta E\left[h_1^j\right] = 0.$$

The last result is the consequence of the envelope theorem. In particular, $\min_\gamma$ will choose $\gamma_j = 0$ if $h_0^j + \beta E\left[h_1^j\right] > 0$. Recursive Contracts allow us to transform a problem with forward-looking constraints into a Bellman problem with prior-period Lagrange multipliers as additional state variables. When facing a constraint, the actor with commitment can change his behavior this period as well as in the future. He balances the adjustment to his decision in whatever way maximizes his welfare. The prior period's Lagrange multipliers represent a constraint's shadow cost to welfare, and by incorporating that cost when making his decision, the agent behaves as he would have committed to behave in the past.

The prior Lagrange multipliers are often described as representing prior promises. In the standard optimization, an agent with commitment makes state-contingent policies and promises to follow through with them indefinitely. When the recursive-contract agents incorporate the prior-period Lagrange multipliers, the recursive policies match the state-contingent policies. Hence, incorporating the prior-period multipliers is described as "respecting prior promises" or even "following through on prior promises." Under that interpretation, the state-contingent policies are the true plans or promises, and Recursive Contracts are just a numerical algorithm for us to calculate them.

There are two problems with this interpretation. First, when using Recursive Contracts is the only way for us to solve the problem, it is better to interpret agent as treating the problem recursively as well. If our best algorithms can only solve it recursively, it is inaccurate to claim the agent is following through on a state-contingent plan he made in the past. Instead, he is also following through on a recursive algorithm. Second, we should take seriously the idea that a timeless agent this period is concerned with how his decision process now would have affected himself in previous periods. It is a backward bargain through time. He wants himself at $t+1$ to consider the effect on t, therefore he considers the effect on $t-1$. This agent exactly follows Woodford's advice, "Rather than doing one thing now but promising to behave differently in the future, one should follow a time-invariant policy that is of the kind that one would always wish *to have been expected* to follow."

3.4.1 General Intermediate Commitment

I now derive the extension of Recursive Contracts for an arbitrary internalization of prior constraints, $\mathcal{T}$

$$V(x,\gamma_{-1}) = \min_{\gamma \in R^l_+} \max_a F(x,a) + \mathcal{T}(x,\gamma_{-1},a) + \gamma h_0(x,a) + \\ + V^{\mathcal{T},adj}(x,\gamma_{-1},\gamma) + \beta E[V(x',\gamma)|x,a]$$

Inside the maximization, $\mathcal{T}$ represents internalization of the timeless perspective. $V^{\mathcal{T},adj}$ is the adjustment to next period's value function from this period's perspective. In the equilibrium, it must be the case that

$$V^{\mathcal{T},adj}(x,\gamma_{-1},\gamma) + \beta E[\mathcal{T}(x',\gamma,a')] = \gamma\beta E_t[h_1(x',a')]$$

and

$$\frac{\partial V^{\mathcal{T},adj}(u,\gamma_{-1},\gamma) + \beta E[\mathcal{T}(x',\gamma,a')]}{\partial\gamma} = \beta E_t[h_1(x',a')]$$

It is not possible for all potential $\mathcal{T}$. However, if $\mathcal{T}$ is linear in γ_{-1}, then we can guarantee a result

$$\mathcal{T}(x,\gamma_{-1},a)=\gamma_{-1}f(x,a)$$

$$V^{\mathcal{T},adj}(x,\gamma_{-1},\gamma)=\gamma\left(g(x,\gamma_{-1})-E\left[f(x',a')\,|x,\gamma_{-1}\right]\right)$$

where it must be the case that

$$g(x,\gamma_{-1})=E\left[h_0(x',a')\,|x,\gamma_{-1}\right]$$

The policy choices a' will probably depend on γ, $a^*(x',\gamma)$, so g must represent the reduced form expectation of the optimal choice $\gamma^*(x,\gamma_{-1})$.

3.4.2 General Scaled Commitment

Linear methods are often used to solve the dynamics of the economy under commitment. In particular if the loss function is quadratic, and the system is linear, then it can be formulated as a linear quadratic regulator (LQR) problem from optimal control. The problem then becomes a matter of linear algebra instead of value function iteration. An attractive feature of scaled commitment is that it only requires a small modification of existing commitment solutions to be able to model varying levels of commitment. Specifically, equation (3.6) includes the term $\frac{H'\tau}{\beta}$ instead of the standard $\frac{H'}{\beta}$. That is all that is needed to introduce scaling parameter τ into a standard LQR commitment model.

This section will follow the derivation in Svensson (2010). Suppose the system is as follows:

$$\begin{bmatrix} I & 0 \\ 0 & H \end{bmatrix}\begin{bmatrix} X_{t+1} \\ E_t x_{t+1} \end{bmatrix}=\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\begin{bmatrix} X_t \\ x_t \end{bmatrix}+\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} i_t+\begin{bmatrix} C \\ 0 \end{bmatrix}\varepsilon_{t+1}. \tag{3.5}$$

X represent predetermined variables, x represent forward-looking variables, i represent instruments, and ε represents zero-mean, iid shocks. Assume A_{22} is invertible, so we have

$$x_t=A_{22}^{-1}\left(HE_t\left[x_{t+1}\right]-A_{21}X_t-B_2 i_t\right).$$

It may seem restrictive that ε_t only affect X_t and not affect x_t. However, one can just add an additional state variable to X_t, and then allow it can impact x_t via A_{21} and A_{22}.

As an example, consider wanting to model the New Keynesian Phillips curve from the first part. We eventually want to capture

$$\pi_t = y_t + \varepsilon_t + \beta E_t\left[\pi_t\right].$$

y_t will be our instrument, π_t must be one of the forward-looking variables. Thus we definite it as follows: $X_t = e_t$, $x_t = \pi_t$, and $i_t = y_t$. (The previous discussion assumed that the shock was iid, which would correspond to $\rho = 0$.)

$$\begin{bmatrix} 1 & 0 \\ 0 & \beta \end{bmatrix} \begin{bmatrix} e_{t+1} \\ \pi_{t+1|t} \end{bmatrix} = \begin{bmatrix} \rho & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} e_t \\ \pi_t \end{bmatrix} + \begin{bmatrix} 0 \\ -\kappa \end{bmatrix} y_t + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \varepsilon_{t+1}.$$

Now assume that the agent is maximizing a quadratic period loss function

$$L_t \equiv -\frac{1}{2} \begin{bmatrix} X_t \\ x_t \\ i_t \end{bmatrix}' \begin{bmatrix} W_{XX} & W_{Xx} & W_{Xi} \\ W_{xX} & W_{xx} & W_{xi} \\ W_{iX} & W_{ix} & W_{ii} \end{bmatrix} \begin{bmatrix} X_t \\ x_t \\ i_t \end{bmatrix}$$

in our case

$$L_t = -\frac{1}{2} \begin{bmatrix} e_t \\ \pi_t \\ y_t \end{bmatrix}' \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \lambda \end{bmatrix} \begin{bmatrix} e_t \\ \pi_t \\ y_t \end{bmatrix}$$

This represents a more complicated system than in the previous section. That is because the upper block was not represented on the previous setup. We need two additional co-state variables to represent the Lagrange multipliers for the two blocks of the system. The first ξ_{t+1} are the Lagrange multipliers on the constraint

$$X_{t+1} = \begin{bmatrix} A_{11} & A_{12} & B_1 \end{bmatrix} \begin{bmatrix} X_t \\ x_t \\ i_t \end{bmatrix} + C\varepsilon_{t+1}.$$

It is denoted with subscript $t+1$ because its value is only known upon realization of ε_{t+1}.

The second co-state is γ_t on the constraint

$$Hx_{t+1|t} = \begin{bmatrix} A_{21} & A_{22} & B_2 \end{bmatrix} \begin{bmatrix} X_t \\ x_t \\ i_t \end{bmatrix}$$

and has subscript t because it is known before the realization of ε_{t+1}.

Agents are maximizing

$$L_t = -\frac{1}{2} \begin{bmatrix} X_t \\ x_t \\ i_t \end{bmatrix}' W \begin{bmatrix} X_t \\ x_t \\ i_t \end{bmatrix}$$

subject to these two constraints. All agents are concerned with future Lagrange multipliers therefore at time t, the effect on utility of

$$\xi_{t+1|t} \left(-A_{11}X_t - A_{12}x_t - B_1 i_t - C\varepsilon_{t+1} \right)$$

as well as the current period multiplier

$$\xi_{t+1}X_{t+1}$$

enters the consideration.

When making choices at time t, all agents also consider the γ_t constraint:

$$\gamma_t \left(Hx_{t+1|t} - \begin{bmatrix} A_{21} & A_{22} & B_2 \end{bmatrix} \begin{bmatrix} X_t \\ x_t \\ i_t \end{bmatrix} \right).$$

Where they differ is how much they incorporate $\frac{1}{\beta}\gamma_{t-1}Hx_t$. Agents with commitment fully incorporate it, and agents with discretion ignore it. We will denote the scaled commitment factor τ as a diagonal matrix whose elements are in $[0, 1]$, with

0 representing discretion and 1 representing commitment. Thus, our scaled commitment agents will incorporate $\frac{1}{\beta}\tau\gamma_{t-1}Hx_t$ when making their choices.

Putting it all together, in period t, an agent maximizes

$$E_t\Bigg[L_t - \xi'_{t+1}\left(\begin{bmatrix}A_{11} & A_{12} & B_1\end{bmatrix}\begin{bmatrix}X_t\\ x_t\\ i_t\end{bmatrix}\right) - \gamma'_t\left(\begin{bmatrix}A_{21} & A_{22} & B_2\end{bmatrix}\begin{bmatrix}X_t\\ x_t\\ i_t\end{bmatrix}\right)$$
$$+\frac{\xi_t}{\beta}X_t + \frac{1}{\beta}\tau\gamma'_{t-1}Hx_t\Bigg]$$

The first order conditions for this maximization of X_t, x_t, and i_t are, respectively

$$0 = \begin{bmatrix}X_t\\ x_t\\ i_t\end{bmatrix}'\begin{bmatrix}W_{XX}\\ W_{xX}\\ W_{iX}\end{bmatrix} - \xi'_{t+1|t}A_{11} - \gamma'_tA_{21} + \frac{1}{\beta}\xi'_t$$

$$0 = \begin{bmatrix}X_t\\ x_t\\ i_t\end{bmatrix}'\begin{bmatrix}W_{Xx}\\ W_{xx}\\ W_{ix}\end{bmatrix} - \xi'_{t+1|t}A_{12} - \gamma'_tA_{22} + \frac{1}{\beta}\tau\gamma'_{t-1}H$$

$$0 = \begin{bmatrix}X_t\\ x_t\\ i_t\end{bmatrix}'\begin{bmatrix}W_{Xi}\\ W_{xi}\\ W_{ii}\end{bmatrix} - \xi'_{t+1|t}B_1 - \gamma'_tB_2$$

Appending columns we get the condition

$$0 = \begin{bmatrix}X_t\\ x_t\\ i_t\end{bmatrix}'W - \begin{bmatrix}\xi_{t+1|t}\\ \gamma_t\end{bmatrix}'\begin{bmatrix}A & B\end{bmatrix} + \frac{1}{\beta}\begin{bmatrix}\xi_t\\ \gamma_{t-1}\end{bmatrix}'\begin{bmatrix}I & 0 & 0\\ 0 & \tau H & 0\end{bmatrix}$$

We can rewrite this as

$$\begin{bmatrix}A & B\end{bmatrix}'\begin{bmatrix}\xi_{t+1|t}\\ \gamma_t\end{bmatrix} = W\begin{bmatrix}X_t\\ x_t\\ i_t\end{bmatrix} + \frac{1}{\beta}\begin{bmatrix}I & 0\\ 0 & H'\tau\\ 0 & 0\end{bmatrix}\begin{bmatrix}\xi_t\\ \gamma_{t-1}\end{bmatrix}$$

Combining this with equation 3.5,

$$\begin{bmatrix} I & 0 & 0 & 0 & 0 \\ 0 & H & 0 & 0 & 0 \\ 0 & 0 & 0 & A'_{11} & A'_{21} \\ 0 & 0 & 0 & A'_{12} & A'_{22} \\ 0 & 0 & 0 & B'_1 & B'_2 \end{bmatrix} \begin{bmatrix} X_{t+1} \\ x_{t+1|t} \\ i_{t+1|t} \\ \xi_{t+1|t} \\ \gamma_t \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & B_1 & 0 & 0 \\ A_{21} & A_{22} & B_2 & 0 & 0 \\ W_{XX} & W_{Xx} & W_{Xi} & \frac{I}{\beta} & 0 \\ W_{xX} & W_{xx} & W_{xi} & 0 & \frac{H'\tau}{\beta} \\ W_{iX} & W_{ix} & W_{ii} & 0 & 0 \end{bmatrix} \begin{bmatrix} X_t \\ x_t \\ i_t \\ \xi_t \\ \gamma_{t-1} \end{bmatrix} + \begin{bmatrix} C \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \varepsilon_{t+1} \tag{3.6}$$

This is identical to the derivation in Svensson (2010), but for the $\frac{H'\tau}{\beta}$ term. In the relevant part of his analysis, commitment is assumed, so it is just $\frac{H'}{\beta}$.

Still following the other derivation, we will rearrange the matrix into blocks, so that the predetermined variables, X_{t+1} and γ_t, are together, and the remaining ones are non-predetermined.

$$\begin{bmatrix} I & 0 & 0 & 0 & 0 \\ 0 & A'_{22} & 0 & 0 & A'_{12} \\ 0 & 0 & H & 0 & 0 \\ 0 & A'_{21} & 0 & 0 & A'_{11} \\ 0 & B'_2 & 0 & 0 & B'_1 \end{bmatrix} \begin{bmatrix} X_{t+1} \\ \gamma_t \\ x_{t+1|t} \\ i_{t+1|t} \\ \xi_{t+1|t} \end{bmatrix} = \begin{bmatrix} A_{11} & 0 & A_{12} & B_1 & 0 \\ W_{xX} & \frac{H'\tau}{\beta} & W_{xx} & W_{xi} & 0 \\ A_{21} & 0 & A_{22} & B_2 & 0 \\ W_{XX} & 0 & W_{Xx} & W_{Xi} & \frac{I}{\beta} \\ W_{iX} & 0 & W_{ix} & W_{ii} & 0 \end{bmatrix} \begin{bmatrix} X_t \\ \gamma_{t-1} \\ x_t \\ i_t \\ \xi_t \end{bmatrix} + \begin{bmatrix} C \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \varepsilon_{t+1}$$

Given this setup, there is a unique solution of the system if the Schur decomposition has the same number of unstable eigenvalues as the number of non-predetermined variables: x_t, i_t, ξ_t. Again see Svensson (2010) for more details.

The solution method yields

$$x_t = F_x \begin{bmatrix} X_t \\ \gamma_{t-1} \end{bmatrix}$$

$$i_t = F_i \begin{bmatrix} X_t \\ \gamma_{t-1} \end{bmatrix}$$

$$\begin{bmatrix} X_{t+1} \\ \gamma_t \end{bmatrix} = M \begin{bmatrix} X_t \\ \gamma_{t-1} \end{bmatrix} + \tilde{C}\varepsilon_{t+1}$$

where the matrices F_x, F_i, and M are independent of C.

The takeaway is that scaled commitment can be a drop-in addition to a common existing solution to forward-looking problems. It allows commitment to be on a spectrum between discretion and commitment.

3.5 Conclusion

This paper presents a new approach to considering intermediate behavior between commitment and discretion. Like other agents who face forward looking constraints, central bankers face the temptation to ignore the past. They may always face it with respect to the inflation bias, and they additionally face it after a period at the zero lower bound. I use one application of the approach to describe how a central bank partially incorporates prior constraints into its decision this period, yielding intermediate values for smoothing shocks and inflation bias. Further research will explore the consequences of other limited internalization of prior constraints, as well as estimating the values from the data.

4

Conclusion

This dissertation discusses two contributions to the macroeconomic theory of monetary policy. Chapter 2 proves the equivalence of three formulations of the problem facing a policymaker with commitment and an informational advantage over the private sector. In that context, the private-sector updating also becomes a control variable to the policymaker. In a New Keynesian central bank model, I also demonstrate how the final, recursive formulation is amenable to finding a steady state. I compare the results for a central bank facing full information and commitment, and one facing an informational advantage under discretion.

Chapter 3 develops an alternative framing of commitment. The standard approach of commitment as the capacity to bind future actions ends up with promises being all-or-nothing, either followed through or ignored. Building off of a micro-theoretic model of temptation and self-control costs, I derive a novel intermediate form of commitment. I then show implications of the intermediate form for a New Keynesian monetary model with and without the zero lower bound constraint.

Appendix A

Appendices for Chapter 2

A.1 Recursive constraints alternative formation

The recursive constraints described starting on page 34 use convenience notation such as $a_{c|p}$, $x_{c|p}$, $a_{c|c}$, $x_{c|c}$, etc. This section precisely defines how those values are calculated based on the augmented state and policy matrices. Recall,

$$y_c \equiv \begin{bmatrix} x_c \\ x_{c|p} \\ a_{c|p} \end{bmatrix} \sim N\left(0, \Sigma^y\right)$$

with

$$\begin{bmatrix} x_{c|p} \\ a_{c|p} \end{bmatrix} = E\left\{ \begin{bmatrix} x_c \\ a_c \end{bmatrix} |I_p^{ps} \right\}$$

Start by defining a variety of convenience matrices for intermediate values that depend on y_c

$$
\begin{aligned}
e_c^x &\equiv \begin{bmatrix} I & 0 & 0 \end{bmatrix} \\
x_c &= e_c^x y_c \\
e_{c|p}^x &\equiv \begin{bmatrix} 0 & I & 0 \end{bmatrix} \\
x_{c|p} &= e_{c|p}^x y_c \\
e_{c|p}^{x|ps} &\equiv e_c^x - e_{c|p}^x \\
x_c - x_{c|p} &= e_{c|p}^{x|ps} y_c \\
e_{c|p}^a &\equiv \begin{bmatrix} 0 & 0 & I \end{bmatrix} \\
a_{c|p} &= e_{c|p}^a y_c
\end{aligned}
$$

Observe that equation (2.13) can be rewritten as

$$
\begin{aligned}
a_c &= a_{c|p} + G^e\left(\Sigma^y\right)\left(x_c - x_{c|p}\right) + \eta_c \\
&= \left(e_{c|p}^a + G^e\left(\Sigma^y\right) e_{c|p}^{x|ps}\right) y_c + \eta_c
\end{aligned}
$$

For convenience define G_c^{xa}

$$
G_c^{xa} \equiv \begin{bmatrix} e^x & 0 \\ e_{c|p}^a + G^e\left(\Sigma^y\right) e_{c|p}^{x|ps} & I_{N_a} \end{bmatrix} \tag{A.1}
$$

$$
\begin{bmatrix} x_c \\ a_c \end{bmatrix} = G_c^{xa} \begin{bmatrix} y_c \\ \eta_c \end{bmatrix}
$$

remembering that G_c^{xa} is a function of Σ^y via G^e. The fully detailed version of equations (2.14) and (2.15), which define L_c and x_n respectively are

$$
L_c = \begin{bmatrix} y_c \\ \eta_c \end{bmatrix}^T \left(G_c^{xa}\right)^T L G_c^{xa} \begin{bmatrix} y_c \\ \eta_c \end{bmatrix}
$$

$$
x_n = \begin{bmatrix} A G_c^{xa} & B \end{bmatrix} \begin{bmatrix} y_c \\ \eta_c \\ w_n \end{bmatrix}
$$

Tracking the private sector's information can be similarly calculated for versions of (2.16) and (2.17),

$$z_c = CG_c^{xa} \begin{bmatrix} y_c \\ \eta_c \end{bmatrix}$$

$$I_c^{ps} = \begin{bmatrix} x_{c|p} \\ a_{c|p} \\ z_c \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} e_{c|p}^{x} & 0 \\ e_{c|p}^{a} & 0 \end{bmatrix} \\ CG_c^{xa} \end{bmatrix} \begin{bmatrix} y_c \\ \eta_c \end{bmatrix}$$

Now to define the private sector belief updates. To define this precisely in terms of y_c and η_c we can use the following facts

$$\begin{bmatrix} y_{c|p} \\ \eta_{c|p} \end{bmatrix} = \begin{bmatrix} e_{c|p}^{x} \\ e_{c|p}^{x} \\ e_{c|p}^{a} \\ 0 \end{bmatrix} y_c = \begin{bmatrix} 0 & I & 0 \\ 0 & I & 0 \\ 0 & 0 & I \\ 0 & 0 & 0 \end{bmatrix} y_c$$

Define the convenience matrix $e_{c|p}^{y\eta|ps}$, where I'm using the notation from the paper of superscript $|ps$ indicating a private-sector prediction error

$$e_{c|p}^{y\eta|ps} \equiv I - \begin{bmatrix} e_{c|p}^{x} & 0 \\ e_{c|p}^{x} & 0 \\ e_{c|p}^{a} & 0 \\ 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} y_c \\ \eta_c \end{bmatrix} - \begin{bmatrix} y_{c|p} \\ \eta_{c|p} \end{bmatrix} = e_{c|p}^{y\eta|ps} \begin{bmatrix} y_c \\ \eta_c \end{bmatrix}$$

Recall that the vector y_c, η_c has the following distribution

$$\begin{bmatrix} y_c \\ \eta_c \end{bmatrix} \sim N\left(0, \begin{bmatrix} \Sigma^y & 0 \\ 0 & \Sigma^\eta\left(\Sigma^y\right) \end{bmatrix}\right)$$

Therefore the following distributions hold, defining convenience matrix $G_{c|p}^{xa|ps}$

$$G_{c|p}^{xa|ps} \equiv G_c^{xa} e_{c|p}^{y\eta|ps}$$

$$\begin{bmatrix} x_c \\ a_c \end{bmatrix} - \begin{bmatrix} x_{c|p} \\ a_{c|p} \end{bmatrix} = G_{c|p}^{xa|ps} \begin{bmatrix} y_c \\ \eta_c \end{bmatrix}$$

$$\Sigma_{c|p}^{xa|ps} = \mathrm{Var}\left(\begin{bmatrix} x_c \\ a_c \end{bmatrix} - \begin{bmatrix} x_{c|p} \\ a_{c|p} \end{bmatrix} \right) = G_{c|p}^{xa|ps} \begin{bmatrix} \Sigma^y & 0 \\ 0 & \Sigma^\eta \left(\Sigma^y\right) \end{bmatrix} \left(G_{c|p}^{xa|ps} \right)^T$$

Per the derivation in appendix A.2, an optimal Kalman gain for z_c is

$$z_c - z_{c|p} = C \left(\begin{bmatrix} x_c \\ a_c \end{bmatrix} - \begin{bmatrix} x_{c|p} \\ a_{c|p} \end{bmatrix} \right)$$

$$K^{xa} \equiv \Sigma_{c|p}^{xa|ps} C^T \left(C \Sigma_{c|p}^{xa|ps} C^T \right)^+ \tag{A.2}$$

with + representing the Moore-Penrose pseudoinverse. Again note that K^{xa} depends on Σ^y, G^e, and Σ^η. Now for the precise definition of equation (2.19), and define convenience matrix $P_{c|c}^{xa}$

$$P_{c|c}^{xa} \equiv \begin{bmatrix} e_{c|p}^x & 0 \\ e_{c|p}^a & 0 \end{bmatrix} + K^{xa} C G_{c|p}^{xa|ps} \tag{A.3}$$

$$\begin{bmatrix} x_{c|c} \\ a_{c|c} \end{bmatrix} = \begin{bmatrix} x_{c|p} \\ a_{c|p} \end{bmatrix} + K^{xa} \left(z_c - C \begin{bmatrix} x_{c|p} \\ a_{c|p} \end{bmatrix} \right)$$

$$= P_{c|c}^{xa} \begin{bmatrix} y_c \\ \eta_c \end{bmatrix}$$

The exact value of y_n in equation (2.21) can be calculated as follows. Recall,

$$\begin{bmatrix} x_{n|c} \\ a_{n|c} \end{bmatrix} = \begin{bmatrix} A \\ G^c \left(\Sigma^y\right) \end{bmatrix} \begin{bmatrix} x_{c|c} \\ a_{c|c} \end{bmatrix}$$

$$= \begin{bmatrix} A \\ G^c \left(\Sigma^y\right) \end{bmatrix} P_{c|c}^{xa} \begin{bmatrix} y_c \\ \eta_c \end{bmatrix}$$

therefore

$$y_n \equiv \begin{bmatrix} x_n \\ x_{n|c} \\ a_{n|c} \end{bmatrix} = \begin{bmatrix} AG_c^{xa} & B \\ AP_{c|c}^{xa} & 0 \\ G^c\left(\Sigma^y\right) P_{c|c}^{xa} & 0 \end{bmatrix} \begin{bmatrix} y_c \\ \eta_c \\ w_n \end{bmatrix} \tag{A.4}$$

$$\Sigma_n^y = \begin{bmatrix} AG_c^{xa} & B \\ AP_{c|c}^{xa} & 0 \\ G^c\left(\Sigma^y\right) P_{c|c}^{xa} & 0 \end{bmatrix} \begin{bmatrix} \Sigma^y & 0 & 0 \\ 0 & \Sigma^\eta\left(\Sigma^y\right) & 0 \\ 0 & 0 & I_{N_w} \end{bmatrix} \left(\begin{bmatrix} AG_c^{xa} & B \\ AP_{c|c}^{xa} & 0 \\ G^c\left(\Sigma^y\right) P_{c|c}^{xa} & 0 \end{bmatrix} \right)^T$$

Finally the precise formula for checking wither the constraint in equation (2.22) holds. Define a N_μ element random variable b_c to be equal to the constraint

$$P_c^b \equiv DG_c^{xa} + J \begin{bmatrix} A \\ G^c\left(\Sigma^y\right) \end{bmatrix} P_{c|c}^{xa}$$

$$b_c \equiv P_c^b \begin{bmatrix} y_c \\ \eta_c \end{bmatrix} \tag{A.5}$$

$$= D \begin{bmatrix} x_c \\ a_c \end{bmatrix} + J \begin{bmatrix} x_{n|c} \\ a_{n|c} \end{bmatrix}$$

$$\text{Var}\left(b_c\right) = P_c^b \begin{bmatrix} \Sigma^y & 0 \\ 0 & \Sigma^\eta\left(\Sigma^y\right) \end{bmatrix} \left(P_c^b\right)^T$$

Lemma 24. *Let b_c be defined in equation (A.5). Forward Looking Constraint (2.22) holds for all* $\begin{bmatrix} y_c \\ \eta_c \end{bmatrix} \in supp\left(N\left(0, \begin{bmatrix} \Sigma^y & 0 \\ 0 & \Sigma^\eta\left(\Sigma^y\right) \end{bmatrix}\right)\right)$ *iff* $tr\left(Var\left(b_c\right)\right) = 0$.

Proof. $\Longrightarrow$, by contrapositive: $tr\left(\text{Var}\left(b_c\right)\right) \neq 0$ implies

$$\exists \begin{bmatrix} y_c \\ \eta_c \end{bmatrix} \in \text{supp}\left(N\left(0, \begin{bmatrix} \Sigma^y & 0 \\ 0 & \Sigma^\eta\left(\Sigma^y\right) \end{bmatrix}\right)\right)$$

such that $0 \neq P_c^b \begin{bmatrix} y_c \\ \eta_c \end{bmatrix} = D \begin{bmatrix} x_c \\ a_c \end{bmatrix} + J \begin{bmatrix} x_{n|c} \\ a_{n|c} \end{bmatrix}$

$\Longleftarrow$,by contrapositive: Assume there exists

$$\exists \begin{bmatrix} y_c \\ \eta_c \end{bmatrix} \in \text{supp}\left(N\left(0, \begin{bmatrix} \Sigma^y & 0 \\ 0 & \Sigma^\eta\left(\Sigma^y\right) \end{bmatrix}\right)\right)$$

such that $P_c^b \begin{bmatrix} y_c \\ \eta_c \end{bmatrix} = b_c \neq 0$. There exists a $\epsilon > 0$ such that

$$\begin{bmatrix} \tilde{y}_c \\ \tilde{\eta}_c \end{bmatrix} \in B\left(\begin{bmatrix} y_c \\ \eta_c \end{bmatrix}, \epsilon\right) \cap \text{supp}\left(N\left(0, \begin{bmatrix} \Sigma^y & 0 \\ 0 & \Sigma^\eta\left(\Sigma^y\right) \end{bmatrix}\right)\right) \implies 0 \neq P_c^b \begin{bmatrix} \tilde{y}_c \\ \tilde{\eta}_c \end{bmatrix}$$

where $B\left(\begin{bmatrix} y_c \\ \eta_c \end{bmatrix}, \epsilon\right)$ is the ϵ-ball around $\begin{bmatrix} y_c \\ \eta_c \end{bmatrix}$. Because $b_c \neq 0$ for a subset of positive measure of $\text{supp}\left(N\left(0, \begin{bmatrix} \Sigma^y & 0 \\ 0 & \Sigma^\eta\left(\Sigma^y\right) \end{bmatrix}\right)\right)$, some element of b_c has strictly positive variance, so the trace of its variance is strictly positive, $tr\left(\text{Var}\left(b_c\right)\right) > 0$. □

By the above lemma, we can determine from Σ^y, G^e, G^c, and Σ^η whether the constraint is met for all possible $\begin{bmatrix} y_c \\ \eta_c \end{bmatrix}$.

A.2 Generalized Kalman updating with the Moore–Penrose pseudoinverse

Coming into period t, let the observer have some uncertainty in state s_t,

$$\Sigma_{t|t-1}^{s|obs} = \text{Var}\left(s_t - s_{t|t-1}\right)$$

A standard Kalman update has the form

$$s_{t|t} = s_{t|t-1} + K_t\left(z_t - z_{t|t-1}\right)$$

for the state s_t, and some signal $z_t = C_t s_t$. If $\Sigma_{t|t-1}^{z|obs} = \text{Var}\left(z_t - z_{t|t-1}\right) = C_t \Sigma_{t|t-1}^{s|obs} C_t^T$ has full rank, the optimal K has the well known form based on the inverse of $\Sigma_{t|t-1}^{z|obs}$.

$$K_t = \Sigma_{t|t-1}^{s} C_t^T \left(\Sigma_{t|t-1}^{z|obs}\right)^{-1}$$

The Kalman filter was originally used to model real-world measurements, so there was some assumed measurement error, $v_t \sim N\left(0, \Sigma^v\right)$, $z_t = C_t s_t + v_t$. If there is some independent error for every dimension of z_t, then $\Sigma_{t|t-1}^{z|obs}$ would always have full rank.

For the purposes of theoretical macroeconomics however, we should entertain the possibility that a fully-informed policymaker could execute a_t perfectly without error. In those cases and under some model specifications, the policymaker may *choose* to make $\Sigma^{z|obs}_{t|t-1}$ have sub-rank. Alternatively, $s_t = s_{t|t-1}$, there is no new information in s_t, and the private-sector is perfectly informed coming into the period $\Sigma^{z|obs}_{t|t-1}$ will be 0.

Here I derive a more general equation for an optimal Kalman gain.[1]

$$
\begin{aligned}
\Sigma^{s|obs}_{t|t} &= \mathrm{Cov}\left(s_t - s_{t|t}\right) \\
&= \mathrm{Cov}\left(s_t - \left(s_{t|t-1} + K_t C_t \left(s_t - s_{t|t-1}\right)\right)\right) \\
&= \mathrm{Cov}\left(\left(I - K_t C_t\right)\left(s_t - s_{t|t-1}\right)\right) \\
&= \left(I - K_t C_t\right) \Sigma^{s|obs}_{t|t-1} \left(I - K_t C_t\right)^T \\
&= \Sigma^{s}_{t|t-1} - K_t C_t \Sigma^{s|obs}_{t|t-1} - \Sigma^{s|obs}_{t|t-1} C_t^T K_t^T + K_t C_t \Sigma^{s|obs}_{t|t-1} C_t^T K_t^T
\end{aligned}
$$

Minimizing $E\left\{\left|s_t - s_{t|t}\right|^2\right\}$ is equivalent to minimizing $\mathrm{tr}\left(\Sigma^{s|obs}_{t|t}\right)$. We can take the derivative to get that any optimal K_t has the property that

$$
\frac{\partial \mathrm{tr}\left(\Sigma^{s|obs}_{t|t}\right)}{\partial K_t} = -2\left(C_t \Sigma^{s|obs}_{t|t-1}\right)^T + 2 K_t \Sigma^{s|obs}_{t|t-1} = 0
$$

rearranging

$$
K_t \Sigma^{z|obs}_{t|t-1} = \Sigma^{s|obs}_{t|t-1} C_t^T
$$

The problem arises if $\Sigma^{z|obs}_{t|t-1}$ doesn't have full rank, so I cannot right multiply by its inverse. In such a case there is more than one optimal K_t. I now show steps to solve for a specific one, using the Moore–Penrose pseudoinverse.

[1] The steps below are similar to those building to equation (4.2.16) in Brown and Hwang (2012) but without their noise term v_t.

Perform an eigenvalue decomposition, $\Sigma_{t|t-1}^{z|obs} = Q\Lambda Q^T$, so that the eigenvalues in Λ are in increasing order of absolute value. If Nullity $\left(\Sigma_{t|t-1}^{z|obs}\right) = n > 0$, then the first n columns of Λ are zero. Right multiply the equation above by Q,

$$K_t Q\Lambda = \Sigma_{t|t-1}^{s|obs} C_t^T Q$$

Thus the first n columns of $\Sigma_{t|t-1}^{s} C_t^T Q$ are also zero. Use the pseudoinverse of Λ

$$\left(\Lambda^{+}\right)_{ii} = \begin{cases} 0 & \lambda_{ii} = 0 \\ 1/\lambda_{ii} & \lambda_{ii} \neq 0 \end{cases}$$

which has the property that

$$\Lambda\Lambda^{+} = \begin{bmatrix} 0 & 0 \\ 0 & I_{N_z - n} \end{bmatrix}$$

Define an optimal K_t^* such that

$$K_t^* Q = \Sigma_{t|t-1}^{s|obs} C_t^T Q \Lambda^{+}$$

and then,

$$\begin{aligned} K_t^* &= \Sigma_{t|t-1}^{s|obs} C_t^T Q \Lambda^{+} Q^T \\ &= \Sigma_{t|t-1}^{s|obs} C_t^T \left(\Sigma_{t|t-1}^{z}\right)^{+} \end{aligned}$$

where I've a property of the more general pseudoinverse, $\left(\Sigma_{t|t-1}^{z}\right)^{+} = Q\Lambda^{+}Q^T$.

A.2.1 Examples

First consider the trivial example where $\Sigma_{t|t-1}^{z|obs} = 0$, implying $z_t = z_{t|t-1}$. Any K is optimal, as it must be the case that $C_t \Sigma_{t|t-1}^{s|obs} = 0$. Thus, no matter the K, $\Sigma_{t|t}^{s|obs} = \Sigma_{t|t-1}^{s|obs}$. The pseudoinverse gives us the reasonable, $K = \Sigma_{t|t-1}^{s|obs} C_t^T (0)^{+} = 0$.

Now consider a somewhat informative, but sub-rank $\left(z_t - z_{t|t-1}\right)$. Let there be two uncorrelated states

$$s_t = \begin{bmatrix} a \\ b \end{bmatrix} \sim N\left(0, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

$$s_{t|t-1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

but one of them gets observed twice.

$$z_t = \begin{bmatrix} a \\ a \end{bmatrix} = C \begin{bmatrix} a \\ b \end{bmatrix}$$

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

then

$$\Sigma^{z|obs}_{t|t-1} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$\left(\Sigma^{z|obs}_{t|t-1}\right)^+ = \begin{bmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{bmatrix}$$

$$K_t^* = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{bmatrix}$$

$$= \begin{bmatrix} 1/2 & 1/2 \\ 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} a_{t|t} \\ b_{t|t} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + K_t^* z_t$$

$$a_{t|t} = \frac{1}{2}\left(z_{t1} + z_{t2}\right)$$

The proposed K_t^* updates the prediction of a_t as the average of the two parts of the signal, z_{t1}, z_{t2}. But any $K_t = \begin{bmatrix} c & 1-c \\ 0 & 0 \end{bmatrix}$ will have the same performance. This comes from the fact that $\Sigma^{z|obs}_{t|t-1}$ has an eigenvalue of 0 for eigenvector $\begin{bmatrix} 1 & -1 \end{bmatrix}^T$, there is no variation in that dimension. So changing the attribution of the signal in that dimension is costless to the final uncertainty in $\Sigma^{s|obs}_{t|t}$.

A.3 Using the pseudoinverse for to match covariance

Consider $y = Ax + \varepsilon$, with $x \sim N(0, \Sigma_x)$. We wish to choose A and Σ_ε so that jointly

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim N\left(0, \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{bmatrix}\right)$$

Let $n = \text{nullity}(\Sigma_x)$. Use an eigenvalue decomposition with an orthonormal basis so that $\Sigma_x = Q\Lambda Q^T$ and the first n eigenvalues in Λ are 0.

Define a rotated $\tilde{x} = Q^T x$, $\text{Var}(\tilde{x}) = \Lambda$, so $j \leq n \implies \tilde{x}_j = 0$. Now consider the covariance of any possible y and $\tilde{x}$

$$\begin{aligned} \text{Cov}(y, \tilde{x}) &= E\left\{y\tilde{x}^T\right\} \\ &= \text{Cov}\left(y, Q^T x\right) \\ &= \Sigma_{yx} Q \end{aligned}$$

But we know that $j \leq n \implies \tilde{x}_j = 0$, so it must be the case that

$$j \leq n \implies (\Sigma_{yx} Q)_{\cdot j} = 0$$

otherwise the desired Σ_{yx} is impossible. Assuming then that $(\Sigma_{yx} Q)_{\cdot j} = 0$, consider this suggestion: $A = \Sigma_{yx}\Sigma_x^+$.

$$\begin{aligned} \text{Cov}(Ax, x) &= \Sigma_{yx}\Sigma_x^+\Sigma_x \\ &= \Sigma_{yx} Q\Lambda^+ Q^T Q\Lambda Q^T \\ &= \Sigma_{yx} Q\Lambda^+\Lambda Q^T \end{aligned}$$

$$\left(\Lambda^+\Lambda\right)_{ii} = \begin{cases} 0 & i \leq n \\ 1 & i > n \end{cases}$$

Because the first n columns of $\Sigma_{yx} Q$ are 0, $\Sigma_{yx} Q\Lambda^+\Lambda = \Sigma_{yx} Q$, and $\text{Cov}(Ax, x) = \Sigma_{yx}$.

For the desired Σ_y to be feasible, it must be the case that $\Sigma_y - \Sigma_{yx}(\Sigma_x)^+\Sigma_{xy}$ is positive semi-definite. Assuming it is, let $\Sigma_\varepsilon = \Sigma_y - \Sigma_{yx}(\Sigma_x)^+\Sigma_{xy}$, and then $\begin{bmatrix} x \\ y \end{bmatrix}$ has the desired covariance.

A.4 Examples of Challenging Full History Sequences

Below are two examples of finite FHS that demonstrate some of the features in the proof. The first shows why η may be necessary for a CES in order to match variances of an FHS. The second shows how an FHS without the Span Property could still be optimal, and how proposition 13 would weakly improve it.

A.4.1 Showing the necessity of η

This example provides a FHS that doesn't use η, but where a CES requires η to match its covariances.

Consider this simple model

$$x_0 \sim N(0, 1)$$

$$z_t = a_t$$

$$x_{t+1} = w_{t+1}$$

Now consider the FHS,

$$a_0 = 0$$

$$a_1 = x_0$$

This almost certainly going to be suboptimal, for interesting loss functions. But proposition 16 says that any FHS with the Span Property can be matched. This plan has the Span Property,

$$z_0 = 0$$

$$\begin{bmatrix} x_{0|0} \\ a_{0|0} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$a_{1|0} = 0$$

Now consider the possible (x_1, a_1) for the CES without η

$$a_1^{CE} = G_0^C \begin{bmatrix} 0 \\ 0 \end{bmatrix} + G_1^e (x_1 - 0) = G_1^e x_1$$

Therefore, its variance must be

$$\text{Var}\left(\begin{bmatrix} x_1 \\ a_1^{CE} \end{bmatrix}\right) = \begin{bmatrix} 1 & G_1^e \\ G_1^e & (G_1^e)^2 \end{bmatrix}$$

The variance for the FHS is

$$\text{Var}\left(\begin{bmatrix} x_1 \\ a_1 \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and cannot be matched for any choice of G_1^e. To match this FHS, the CES must use

$$a_1^{CE*} = \eta_1$$

$$\eta_1 \sim N(0,1)$$

This shows that there exist FHS with the Span Property with $\Sigma_t^\eta = 0$, but whose matching CES must use $\Sigma_t^{\tilde{\eta}} \neq 0$ in order to match variances of (y_t, a_t). The problem arises because the FHS has access to variation outside of $\left(x_t, x_{t-1|t-1}, a_{t-1|t-1}\right)$ which are the random variables from the model available to the CES. Therefore, we must give the CES access to $\Sigma_t^{\tilde{\eta}}$, in order to match all possible covariances (y_t, a_t). Once the CES is using η_t, the FHS must have access as well so that the equivalence goes both ways.

A.4.2 An optimal FHS without the Span Property

This shows a (somewhat trivial) FHS without the Span Property at time 1. L is constructed so that the extra variation in $a_{2|1}$ is irrelevant to losses, and the FHS is still optimal. Finally, I show how proposition 13 would modify the sequence, so that the new FHS has the Span Property at time 1. (I omit η_t as it is not used in this example.)

$$x_0 \sim N(0,1)$$

$$z_t = x_t$$

$$x_{t+1} = w_{t+1}$$

$$L_t = a_{t,1}^2$$

The Forward-Looking Constraint is the simple

$$0 = \begin{bmatrix}1 & -1 & 0\end{bmatrix} \begin{bmatrix} x_t \\ a_t \end{bmatrix} + 0 \begin{bmatrix} x_{t+1|t} \\ a_{t+1|t} \end{bmatrix}$$

$$= x_t - a_{t,1}$$

so it must be the case that $a_{t,1} = x_t$ for all t, and all FHS in $\mathcal{FH}$ have the same losses.

Now consider the following FHS for $t \in \{0, 1, 2\}$,

$$a_0 = \begin{bmatrix} x_0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} x_0$$

$$a_1 = \begin{bmatrix} w_1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ w_1 \end{bmatrix}$$

$$a_2 = \begin{bmatrix} w_2 \\ x_0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ w_1 \\ w_2 \end{bmatrix}$$

we can see that at time 1,

$$I_1^{ps} = \{x_0, w_1\}$$

$$\begin{bmatrix} x_{1|1} \\ a_{1|1} \end{bmatrix} = \begin{bmatrix} x_1 \\ a_1 \end{bmatrix} = \begin{bmatrix} w_1 \\ w_1 \\ 0 \end{bmatrix}$$

$$a_{2|1} = \begin{bmatrix} 0 \\ x_0 \end{bmatrix}$$

The FHS does not have the Span Property at time 1, because $a_{2|1} \notin \text{span}\left(x_{1|1}, a_{1|1}\right) = \text{span}\left(\{w_1\}\right)$. As discussed after proof of proposition 13, this additional variation is irrelevant to the Forward-Looking Constraint. This FHS is still optimal, because the extra variance in $a_{2|1}$ does not affect losses.

Following the proof and using definition 11,

$$H_2^{*1} = \left\{h_2 : h_{2|1} = h_2 \wedge x_1 = 0 \wedge a_1 = 0\right\}$$

the first condition says that $w_2 = 0$, the second and third conditions say that $w_1 = 0$. Therefore,

$$H_2^{*1} = \left\{ \begin{bmatrix} x_0 \\ 0 \\ 0 \end{bmatrix} \right\}$$

$$H_2^{*1\perp} = \left\{ \begin{bmatrix} 0 \\ w_1 \\ w_2 \end{bmatrix} \right\}$$

$$P_2^{*1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Applying the steps of the proof to construct the new $\tilde{G}_2$,

$$G_2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

$$\tilde{G}_2 = G_2 \left(I - P_2^*\right)$$

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

Decomposing a_2, $a_2^{*1} = \begin{bmatrix} 0 \\ x_0 \end{bmatrix}$, this is taken to 0, so $\tilde{a}_2^{*1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. In the orthogonal compliment, $a_2^{*1\perp} = \begin{bmatrix} w_2 \\ 0 \end{bmatrix} = \tilde{a}_2^{*1\perp}$. The final result is

$$\tilde{a}_2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \left(I - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right) \begin{bmatrix} x_0 \\ w_1 \\ w_2 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ w_1 \\ w_2 \end{bmatrix}$$

$$= \begin{bmatrix} w_2 \\ 0 \end{bmatrix}$$

$$\tilde{a}_{2|1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

The new FHS has strictly lower covariance and the Span Property at time 1.

Bibliography

Adam, Klaus and Roberto M Billi (2006): Optimal monetary policy under commitment with a zero bound on nominal interest rates. *Journal of Money, Credit, and Banking*, 38(7):1877–1905.

Beckworth, David (2017): Permanent versus temporary monetary base Injections: Implications for past and future Fed Policy. *Journal of Macroeconomics*, 54:110–126.

Bodenstein, Martin; James Hebden; and Ricardo Nunes (2012): Imperfect credibility and the zero lower bound. *Journal of Monetary Economics*, 59(2):135–149.

Brown, Robert Grover and Patrick Y. C. Hwang (2012): *Introduction to Random Signals and Applied Kalman Filtering with Matlab Exercises, 4th Edition*. Wiley.

Cukierman, Alex and Allan H Meltzer (1986): A theory of ambiguity, credibility, and inflation under discretion and asymmetric information. *Econometrica: journal of the econometric society*, pp. 1099–1128.

Debortoli, Davide and Aeimit Lakdawala (2016): How credible is the federal reserve? a structural estimation of policy re-optimizations. *American Economic Journal: Macroeconomics*, 8(3):42–76.

Debortoli, Davide and Ricardo Nunes (2010): Fiscal policy under loose commitment. *Journal of Economic Theory*, 145(3):1005 – 1032.

Debortoli, Davide and Ricardo Nunes (2014): Monetary regime switches and central bank preferences. *Journal of Money, credit and Banking*, 46(8):1591–1626.

Giannoni, Marc P and Michael Woodford (2017): Optimal target criteria for stabilization policy. *Journal of Economic Theory*, 168:55–106.

Gul, Faruk and Wolfgang Pesendorfer (2001): Temptation and self-control. *Econometrica*, 69(6):1403–1435.

Gul, Faruk and Wolfgang Pesendorfer (2004): Self-Control and the Theory of Consumption. *Econometrica*, 72(1):119–158.

Laibson, David (1997): Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2):443–478.

Marcet, Albert and Ramon Marimon (2019): Recursive Contracts. *Econometrica*, 87(5):1589–1631.

Melosi, Leonardo (2016): Signalling effects of monetary policy. *The Review of Economic Studies*, 84(2):853–884.

Mertens, Elmar (2016): Managing beliefs about monetary policy under discretion. *Journal of Money, Credit and Banking*, 48(4):661–698.

Nakamura, Emi and Jón Steinsson (2018): High-frequency identification of monetary non-neutrality: the information effect. *The Quarterly Journal of Economics*, 133(3):1283–1330.

Svensson, Lars EO (2010): Optimization under commitment and discretion, the recursive saddlepoint method, and targeting rules and instrument rules: Lecture notes.

Svensson, Lars EO and Michael Woodford (2003): Indicator variables for optimal policy. *Journal of monetary economics*, 50(3):691–720.

Svensson, Lars EO and Michael Woodford (2004): Indicator variables for optimal policy under asymmetric information. *Journal of Economic Dynamics and Control*, 28(4):661–690.

Tang, Jenny (2015): *Uncertainty and the signaling channel of monetary policy*. Working Papers 15-8, Federal Reserve Bank of Boston Research Department.

Woodford, M. (2003): *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press.

Woodford, Michael (1999): Commentary: How should monetary policy be conducted in an era of price stability? *New Challenges for Monetary Policy*.

Biography

Craig Fratrik graduated *cum laude* from Rice University in 2006, with a B.A. in Computer Science and Math. He then worked as a software engineer for six years. In 2015, he graduated *cum laude* with a J.D. from Harvard Law School. In 2023, he graduated with a Ph.D. in Economics from Duke University.